The capacity for complex thought is one of the traits that distinguish us as humans. Although social psychologists have been concerned with thinking and thought processes from our earliest history (e.g., Allport, 1954; James, 1890), the contemporary social cognition movement catapulted the importance of thought processes to center stage within the field (e.g., Fiske & Taylor, 1991; Moscovitz, 2004). This thinking is sometimes very effortful and deliberate, and at other times it is more simple and automatic (Chaiken, Liberman, & Eagly, 1987; Kruglanski & Thompson, 1999; Petty & Cacioppo, 1986). Furthermore, our thinking is on occasion quite objective, but at other times is contaminated by various biases (see Eagly & Chaiken, 1993; Petty & Wegener, 1998, for reviews).

Although there are many distinctions such as those just noted that can be made about thoughts, in this chapter we are concerned with a distinction between primary and secondary cognition. Primary thoughts involve our initial associations of some object with some attribute, or a projection of some object on some dimension of judgment such as “the flower is red” or “I like the flower” (McGuire & McGuire, 1991). Our thoughts can be directed at any object including other people, the environment, and ourselves. Following a primary thought, people can also generate other thoughts that occur at a second level which involve reflections on the first level thoughts (e.g., “Is that flower really red or pink?” and “I am not sure how much I like that flower”).

Meta-cognition refers to these second order thoughts, or our thoughts about our thoughts or thought processes. Some authors have conceived meta-cognition more broadly as people’s knowledge about their own and others’ mental states, processes, and beliefs (Jost, Kruglanski, & Nelson, 1998; Wright 2002). In the present chapter, however, we will focus mostly on research examining the role of thinking about one’s own thoughts and thought processes since that is the domain in which most of the social meta-cognitive research has been conducted.

The topic of meta-cognition has received considerable theoretical and empirical attention in the past decade, being considered one of the top 100 topics of psychological research (Nelson 1992). Within the field of cognitive psychology, the study of meta-cognition has traditionally focused on how people monitor and control their own mental functioning, especially in the domain of memory (e.g., Koriat and Goldsmith 1996; Nelson and Narens 1994). For example, research has found that the stronger one’s feeling of knowing about a piece of information, the more time one is willing to spend searching for it (e.g., Costermans, Lories, and Ansay 1992). The motivation to complete the search is particularly intense when one has the subjective experience that the information is on the tip of the tongue and thus about to emerge into consciousness (see Yzerbyt, Lories, and Dardenne 1998). Thus, meta-cognitive processes are consequential in guiding further thinking and action.

In addition to cognitive psychology, where the formal study of meta-cognition emerged, the idea that people think about their thoughts and thought processes is prevalent in a number of other psychological domains. For instance evaluating one’s thoughts is critical to some forms of clinical practice. Indeed, the main goal of cognitive-behavior therapy is to get individuals to further think about their maladaptive thoughts with the goal of inducing doubt in them (e.g., Beck & Greenberg, 1994; Ellis, 1962). Meta-cognition also plays a prominent role in the context of consumer psychology (Alba & Hutchinson, 2000; Wright, 2002). For instance, Friestad and Wright (1994) have noted that people’s naïve theories of attitude change play an important role in determining how individuals deal with persuasion attempts (see also,
Before beginning our review of social psychology’s research on meta-cognition, it is useful to discuss the dimensions along which meta-cognitive thought might be organized. In particular, we note that the dimensions of meta-cognitive thought can be organized along many of the same dimensions that have proven useful for understanding primary thoughts, as well as some unique dimensions. Within social psychology, thought coding is particularly prominent in research on attitude change. In this literature, thoughts generated in response to a persuasive message are typically classified into the following categories by judges or the participants themselves: target (e.g., is the thought about the message content or the source, etc.), origin (e.g., does the thought stem from the message content or is it uniquely generated by the message recipient), valence (e.g., is the thought favorable or unfavorable toward the proposal), and number (e.g., are there many or few thoughts; e.g., see Cacioppo, Harkins, & Petty, 1981; Greenwald, 1968; Wright, 1973). Coding thoughts for target, origin, valence, and number has provided a very fruitful approach for understanding some of the psychological processes that underlie attitude formation and change (see Eagly & Chaiken, 1993; Petty, Ostrom, & Brock, 1981). Interestingly, when participants are asked to judge their own thoughts in persuasion research, they are in essence being asked to engage in meta-cognition because they are asked for their thoughts about their thoughts (e.g., how favorable toward the issue is your thought?). There is no presumption in the literature on primary cognition, however, that people necessarily think about their thoughts in this way on their own—only that these post-hoc categorizations are useful for predicting what attitudes people will adopt. Since these categories have already proven effective for classifying primary thoughts, we postulate that they can also serve to understand thoughts at the meta-cognitive level of thinking.

Thus, in the present chapter we use the same categories to describe meta-cognitive thought. In primary cognition, the dimensions often refer to some objective reality that judges or the participants themselves are asked to determine. For example, with respect to the target of the thought, a judge or the participant might be asked if the thought is about the message itself or about the source of the communication (e.g., Chaiken, 1980). With respect to origin, a judge could determine if the thought reflected original thinking or merely restated the message content (e.g., Greenwald, 1968). With respect to valence, a judge or the participant might be asked if the thought reflects some positive or negative reaction to the message (e.g., Petty, Wells, & Brock, 1976). Perhaps attesting to the rather objective nature of these assessments, judges’ and participants’ ratings often correlate quite well (see Cacioppo et al., 1981). As we describe further below, with meta-cognitive thought, the focus is on the individual’s perception of his or her own thought, regardless of its relation to any objective reality. In meta-cognition, the issue is whether the person spontaneously thinks about the target, origin, valence, quantity or some other dimension of his or her thoughts, and whether these second order thoughts are consequential.

In this chapter, we will refer back to these various dimensions of thought as needed in reviewing the empirical literature. The main function of our organizing structure is to facilitate and highlight the similarities between studies coming from diverse domains in social psychology (e.g., research on the self versus attitude change). Our review of the literature is organized into four substantive content areas. Within each area we focus on the types of thoughts people have about their primary thoughts, and the consequences of these thoughts. We also address, when relevant, thoughts people have about their thought processes, and issues for future research. Before reviewing the core areas in which meta-cognition has been applied, we elaborate the dimensions of meta-cognition that are most studied in the literature.

Types of Thoughts about Thoughts

Briefly described, the target dimension of meta-cognition refers not to what the thought is actually about, but what the person perceives it to be about (cf., Higgins, 1998). For example, a person might wonder: Is the thought about a lost person? Is the thought about me? Among other things, this type of meta-cognition can help individuals to classify their thoughts into categories as a first step in marking them for further control and change (e.g., Ellis, 1962). The origin of a thought refers to its source. Where did the thought come from? That is, a person can wonder if the thought is his or her own, or if it merely reflects the statements or sentiments of others (e.g., Greenwald, 1968). People are more likely to act on thoughts that are perceived as connected to or originating from the self (see Wheeler, DeMarree, & Petty, 2005). Perhaps the most commonly studied aspect of thought is its valence. That is, regardless of the target of the thought or its origin, does the person perceive it to reflect something positive or negative with respect to its target? Finally, amount refers not to how many thoughts are actually generated, but to the perceived quantity of thoughts. For example, people might think that they possess very few or many thoughts about a given topic, and as we review shortly, such thought attributions have

---

1 People’s primary thoughts can also be classified into other dimensions such as fluency (e.g., how accessible the thought is), or time (e.g., how recent the thought is; see also, Cornoldi & Vianello, 1992), but these dimensions are used infrequently in coding thoughts.
important implications for social judgment and behavior (e.g., Schwarz, et al., 1991).

In addition to these aspects of thought, two additional dimensions are uniquely meta-cognitive and have achieved the most conceptual and empirical attention. First, one can assess one’s evaluation of a thought. That is, regardless of the perceived target, origin, valence, or number of thoughts generated, people can assess their thoughts as good or bad, desirable or undesirable, appropriate and wanted, or not. When thoughts are unwanted, people might try to suppress them (Wegner, 1984). Or, if the thoughts are seen as inappropriate or bad, people might try to correct for their anticipated impact on judgments or action (Wegener & Petty, 1997).

Second, people can have varying degrees of confidence in their thoughts, ranging from extreme certainty to a high amount of doubt. Thus, two people might have the same thought with respect to a given proposal or target, but one person might have considerably more confidence in that thought than the other. Thoughts held with more confidence have a larger impact on judgments (e.g., Petty, Briñol, & Tormala, 2002). Given the attention that research in social psychology has dedicated to the study of confidence, in the next section we describe this dimension in more detail.

Before turning to confidence, however, it is worth noting that the different categories of meta-cognitive thoughts likely relate to each other. For example, a thought whose origin is perceived as the self might be evaluated more favorably than one with a perceived external origin (e.g., Greenwald & Albert, 1968). Or, thoughts about targets for which we are perceived to have more knowledge than other people might be held with more confidence (Kruglanski et al., 2005). Furthermore, it is possible to consider third order cognition in which people are asked to make one meta-cognitive judgment about another. Thus, after asking about the evaluation of a particular thought, one can ask about the confidence in that evaluation. Or, after asking about the confidence in one’s thought, one can ask whether this confidence is appropriate, and so forth. To our knowledge, no research to date has systematically examined third order cognition.

Meta-cognitive Confidence

A subjective experience that constitutes one of life’s greatest meta-cognitive challenges is the sense of epistemic certainty or uncertainty (e.g., Kruglanski 1980, 1989; Nelson et al. 1998). Although certainty could presumably apply to various aspects of one’s judgment (e.g., am I certain that this is my belief?), social psychological theory has paid special attention to the perceived validity of one’s judgments. For example, in lay-epistemic theory (Kruglanski, 1990), following the generation of a thought or hypothesis, people are said to want to test and validate the thought. Within dual process models of persuasion such as the elaboration likelihood model (Petty & Cacioppo, 1986), and the heuristic-systematic model (Chaiken, Liberman, & Eagly, 1989), people are postulated to seek a certain level of validity, confidence, or certainty in their judgments. When the gap between a person’s current level of judgmental confidence and the desired level is high, they should engage in greater information processing activity (see Chaiken et al., 1989). Underlying these notions is the assumption that people generally seek to hold accurate judgments (Festinger, 1954), that uncertainty is aversive, and that deliberative information processing is often a good way of obtaining accuracy and reducing uncertainty (cf., Sorrentino et al., 1988). In fact, inducing a general feeling of doubt, whether explicit (Tiedens & Linton, 2001) or implicit (Petty, Tormala, Brinol, & Jarvis, 2006), has provoked greater information processing. Indeed, incongruencies of all sorts have increased information processing activity (e.g., Baker & Petty, 1994; Maheswaran & Chaiken, 1991; Smith & Petty, 1995; Ziegler, Diehl, & Ruther, 2002), presumably by inducing some doubt or confusion that might be resolved with thinking.

Thus, meta-cognitive confidence generally refers to a sense of validity regarding one’s thoughts or judgments, though it is possible for certainty to be applied to other aspects of attitude-relevant beliefs (e.g., certainty that the attitude is my own). Studying meta-cognitive confidence is important mainly because confidence affects whether people translate their individual thoughts into more general judgments or evaluations, and whether these judgments in turn are influential in guiding behavior. There is great deal of empirical evidence suggesting that beliefs held with great conviction are a more potent foundation for judgment and behavior than more tentatively held beliefs (e.g. Berger & Mitchell, 1989; Briñol & Petty, 2004; Fazio & Zanna, 1978; Pieters & Verplanken, 1995; Swann & Ely, 1984; Tormala & Petty, 2002).

Logically, one might expect that highly confident thoughts, beliefs, and attitudes determine actions because confidence stems from a truly accurate perception of reality. That is, common sense would suggest that confidence should emerge from objective sources such as the amount and quality of information people have. For example, the more time a person has to see something, the more confident they should be in their identification of it. However, an objective basis to confidence seems to be the exception rather than the rule. For example, when people were asked to predict the behavior or the personality of patients, prosecutors, dating partners,
roommates or strangers, there was no relation or only a modest correlation between confidence and predictive accuracy (e.g., Oskamp, 1965; Deffenbacher, 1984; Dunning, Griffin, Milojkovic, & Ross, 1990; Jacoby, Bjork, & Kelley, 1994; Swann & Gill, 1997; Wells & Murray, 1984). Numerous studies have shown that the confidence with which people hold their beliefs can be affected by different factors that do not necessarily increase the validity of beliefs. For example, the more frequently people think about their thoughts, the more confidence they have in them (e.g., Koriat, Lichtensetein, & Fischhoff, 1980), the more a judgment is repeated, the more confidence is increased (e.g., Shaw, 1996), the more details that are included in a given thought, the more confidence is increased (e.g., Gill, Swann, & Silvera, 1998). In addition to these factors, confidence has been found to be affected by several other situational (e.g., Yzerbyt, Schadron, Leyens, & Rocher, 1994) and personality (Schaefer, Williams, Goodie, & Campbell, 2004) variables.

Research on decision-making has also provided a number of important examples of how peoples’ judgments of confidence can be independent of their accuracy. According to Griffin and Tversky (1992), for instance, confidence judgments require the integration of different kinds of evidence such as the extremity of the available information and the weight or predictive validity of that information. For example, when evaluating a letter of recommendation for a graduate student written by a former teacher, a person may consider how positive the letter is (i.e., extremity) and how credible or knowledgeable the writer is (i.e., weight). According to Tversky and Kahneman (1974), people tend to focus mostly on the extremity of the evidence leading them to underutilize other variables that control predictive validity. As a result, Griffin and Tversky (1992) argue that over-confidence occurs when extremity is high and weight is low, whereas under-confidence takes place when extremity is low and weight is high.

In this research, as in most of the literature on decision making, confidence judgments refer exclusively to the estimation of how likely it is for an answer (e.g., a judgment or a decision) to be correct, and criteria of accuracy are typically available. In social psychological research it is much less common to use objective criteria of accuracy since people’s thoughts often relate to judgments or actions involving people, groups, political views, and so forth. Although one can determine if one’s confidence in the likelihood of a red ball coming out of an urn is accurate, it is not possible to determine if one’s confidence in an attitude toward a presidential candidate is accurate in any objective sense. Yet, a subjective sense of accuracy or confidence has important implications such as determining if you will vote for the candidate based on your attitude. Within the social psychological literature, judgmental confidence has been tied to the extremity of the attributes a target possesses and the certainty that the target does or does not possess those attributes (Griffin & Tversky, 1992; Petty, et. al., 2002), one’s confidence in the likelihood that a target possesses a certain attribute and one’s confidence in the desirability of that attribute (Briñol, Petty, & Tormala, 2004), as well as other sources such as the ease with which information comes to mind (Haddock, Rothman, Reber, & Schwarz, 1999).

Although meta-cognitive ideas have been applied to various topics in social judgment, in the next sections we focus on some of the most heavily researched areas of social psychology in which the role of meta-cognition has been examined. These are: memory and cognitive fluency, attitudes and persuasion, the self and individual differences, and bias correction. In the final section, several key conclusions and general principles of the reviewed work are outlined.

Memory and Cognitive Fluency

Meta-cognition is deeply rooted in the study of human memory—in particular, people’s theories about and perceptions of their memory (e.g., Costermans, Lories, & Ansay, 1992; Strack & Förster, 1998), their knowledge (e.g., Koriat, 1993), and their learning (e.g., Dunlosky & Nelson, 1994). Jacoby and colleagues, for instance, proposed that memory often involves not a literal search for stored information, but rather a series of inferences based on “cognitive feelings” (e.g., Jacoby, Kelley, & Dywan, 1990; see also Clore & Parrot, 1994). One such feeling is familiarity, which people interpret as indicating that something is known, or remembered (e.g., Jacoby, Kelley, Brown, & Jaseckho, 1989). Another such feeling involves the ease or fluency with which information can be retrieved from memory (e.g., Benjamin & Bjork, 1996; Koriat & Levy-Sadot, 1999; Nelson & Narens, 1990). It is this feeling of fluency that has received the most attention in social meta-cognition research. Next, we provide an overview of classic and contemporary findings on fluency. In so doing, we also review the current state of knowledge with respect to 2 very influential effects in social psychological research: ease of retrieval effects and mere exposure effects.

Ease of Retrieval

We begin with the notion that people sometimes base judgments on the subjective experience of fluency with which information comes to mind. In their seminal research on the availability heuristic, Tversky and Kahneman (1973) found that people sometimes perceive events as more likely or common if examples of them come to mind easily. In their now classic work on ease of retrieval, Schwarz and colleagues (1991) expanded upon the earlier research by directly pitting the subjective ease of information retrieval against the actual content of the information retrieved. Schwarz et al. asked participants to rate their assertiveness after recalling 6
versus 12 examples of their own assertive behavior. They found that people viewed themselves as more assertive after retrieving 6 rather than 12 examples. The logic behind this effect was that people based their judgments of assertiveness on the subjective experience of the ease with which assertive behaviors could be retrieved from memory. When it was easy, because only a few behaviors were requested, people concluded that they must be pretty assertive. When it was difficult, because many behaviors were requested, people inferred that they must not be very assertive. What made this finding so intriguing was that it demonstrated that in at least some contexts people forsake the content of accessible information in memory (e.g., having many assertive behaviors activated) and instead base judgments on the subjective experience of memory (e.g., ease or difficulty). Of course, these effects depend on the perceived diagnosticity of the feeling of ease or difficulty (see Schwarz, 1998, for a review). In the Schwarz et al. (1991) studies, for example, ease of retrieval had no impact on judgments of assertiveness when ease was attributed to an external source.

Since the Schwarz et al. (1991) research, the ease of retrieval notion has been applied to a number of domains. In the attitudes domain, for instance, it had been found that the easier it feels for people to generate positive thoughts about an object or issue, the more people like that object or issue (Haddock, Rothman, & Schwarz, 1996; Tormala, Petty, & Briñol, 2002; Wänke, Bless, & Biller, 1996; Wänke, Bohner, & Jurkowitsch, 1997). In addition to attitudes, ease of retrieval can influence a variety of other judgments, such as likelihood estimates (e.g., Hirt, Kardes, & Markman, 2004; Wänke, Schwarz, & Bless, 1995), risk assessments (Grayson & Schwarz, 1999; Rothman & Schwarz, 1998), stereotypes (Dijksterhuis, Macrae, & Haddock, 1999; see also Rothman & Hardin, 1997), attitude certainty (Haddock et al., 1999), judgments of interpersonal closeness (Broemer, 2001), and feelings of self-doubt (Hermann, Leonardelli, & Arkin, 2002). Recent research has shown that ease of retrieval can affect not only explicit, but also implicit measures of attitudes (Gawronski & Bodenhauen, in press). Behavior can also be affected. For example, Keller and Bless (in press) found that when people were asked to recall negative stereotypes about their own group, their performance on an ability test was worse when they recalled few (easy) rather than many (difficult) pieces of information.

Although the basic ease effect appears to be quite robust, there have been divergent findings with respect to moderating and mediating factors. For instance, are ease effects most likely to occur under high or low thinking conditions? Researchers originally assumed that ease effects are heuristic in nature and, thus, most likely to operate when thinking is low (see Schwarz, 1998). Some evidence has been produced that is consistent with this notion (Grayson & Schwarz, 1999; Rothman & Schwarz, 1998; Ruder & Bless, 2003). Other research, however, has pointed to the exact opposite conclusion — that ease effects are more likely to operate under high thinking conditions, when people have the motivation and ability to attend to and interpret their own cognitive experience (Hirt et al., 2004; Tormala et al., 2002; Wänke & Bless, 2000).

This controversy may stem, at least in part, from different perspectives on the mechanism responsible for ease of retrieval effects. The argument for low thought conditions is based on the notion that ease effects stem from availability inferences (Schwarz, 1998). Difficulty in generating favorable arguments for a tax cut, for example, indicates that few favorable arguments exist, so the tax cut is not worth supporting. According to this explanation, ease effects might be viewed as emerging mainly when thinking is low since that is when heuristics are most likely to impact judgments (Chaiken, 1987). Other research (Tormala et al., 2002), however, suggests that ease effects can be mediated by feelings of confidence or validity associated with the particular arguments or thoughts retrieved. The easier it is to generate a list of arguments supporting a tax cut, the more confident people are that those arguments are valid (see also Wänke & Bless, 2000). People have been found to be particularly attuned to thought confidence and validity under high thinking conditions (Petty et al., 2002). Although research has identified multiple mechanisms by which ease effects can occur, each begins with the assumption that people perceive their ease or difficulty in thinking — a meta-cognition. What differs in the accounts is what inferences people make based on this perceived ease.

**Mere Exposure**

A well known phenomena that predates ease of retrieval research is that of mere exposure. The mere exposure effect occurs when attitudes toward stimuli become more favorable as a consequence of repeated exposure to those stimuli (Zajonc, 1968). In one early demonstration of this phenomenon, Kunst-Wilson and Zajonc (1980) presented people with a series of polygon images and found that even when these images could not be consciously recognized, the more often they had been presented, the more they were liked. This effect has now been demonstrated with a wide variety of stimuli such as foreign words, photographs, music, ideographs, and nonsense syllables (see Bornstein, 1989, for a review). Moreover, it has been shown that mere exposure can affect mood, and that this mood can spread to other, related stimuli that were not even presented (Monahan, Murphy, & Zajonc, 2000). To account for mere exposure effects, two general explanations have been proposed: perceptual fluency and perceived familiarity.

A great deal of research suggests that previous or repeated exposure to stimuli can make those stimuli...
easier to process, and that this perceptual fluency enhances subsequent liking. Specifically, the feeling of perceptual fluency, or ease of processing, is thought to be misattributed to a positive evaluation of the stimulus (Bornstein, 1989; Bornstein & D’Agostino, 1994; Jacoby, Kelley, Brown, & Jasechko, 1989). Of importance, though, perceptual fluency can also be attributed to other stimulus dimensions (Mandler, Nakamura, & Van Zandt, 1987). For instance, previously presented names seem more famous (Jacoby, Woloshyn, & Kelley, 1989) and previously presented statements seem more true (Begg, Armour, & Kerr, 1985), even when those statements are explicitly identified as false (Skurnik, Yoon, Park, & Schwarz, 2005). When stimuli already have some meaning, or tend to elicit a dominant response in one direction or another, repeated exposure can accentuate that dominant response (Brickman, Redfield, Harrison, & Crandall, 1971). Repeatedly presenting negative information, for instance, can make that information seem more negative (Cacioppo & Petty, 1989; Grush, 1976). One possible reason for these polarization effects is that one’s positive assessments of positive information seem more valid or plausible as exposure increases, as do one’s negative assessments of negative information (Kruglanski, Freund, & Bar-Tal, 1996).

As an alternative to the perceptual fluency account, some research suggests that mere exposure effects might at least partially derive from feelings of familiarity (see Bornstein & D’Agostino, 1994; Lee, 2001). In fact, Lee (1994, 2001) found that people generally prefer old (familiar) to new (unfamiliar) stimuli, even when the old stimuli have not been repeatedly presented. Furthermore, some research suggests that the feeling of familiarity enhances liking even when the familiarity does not stem from any prior exposure at all. Moreland and Zajone (1982) found that people responded more favorably to faces when those faces felt familiar because they were similar to other ones that had been viewed. The association between familiarity and liking is so strong that stimuli that are already positive are also likely to be perceived as familiar (Corneille, Monin, & Pleyers, 2005; Garcia-Marques, Mackie, Claypool, & Garcia-Marques, 2002; Monin, 2003). As with perceptual fluency, however, some studies have qualified this view, suggesting that the feeling of familiarity stemming from repeated exposure can foster liking or disliking, depending on other contextual factors (e.g., Klinger & Greenwald, 1994; for a similar finding, see Smith, Miller et al., in press).

Although the perceptual fluency and familiarity explanation of mere exposure effects have been pitted against one another in the literature, these mechanisms might ultimately be somewhat intertwined. For example, Whittlesea, Jacoby, and Girard (1990) found that stimuli presented with greater visual clarity were perceived as being more familiar, and more likely to have appeared previously, than stimuli presented with less visual clarity. Furthermore, perceptual fluency stemming from repeated exposure can make a stimulus feel more familiar, and enhance liking in this manner (e.g., Jacoby & Kelley, 1987; Jacoby et al., 1989).

Other Sources of Fluency

In addition to ease of retrieval and mere exposure, there are a number of other sources of processing fluency. The classic perceptual fluency effect—that easy-to-process stimuli are evaluated more favorably than difficult-to-process stimuli—has been produced using a diverse set of experimental manipulations. Line drawings, for example, tend to be liked more and produce more positive affect when they have greater rather than less figure-ground contrast, when they have been presented for a longer rather than shorter amount of time, and when they have been preceded by a similar rather than dissimilar prime (Reber, Winkielman, & Schwarz, 1998; see also Winkielman & Cacioppo, 2001). In addition, words tend to be easier to process and judged as more pleasant when they are embedded in a predictive rather than nonpredictive semantic context (e.g., Whittlesea, 1993).

Other Effects of Fluency

Confidence

In addition to affecting people’s evaluations of stimuli, as noted earlier, processing fluency can also influence feelings of confidence. For example, recall that the ease of generating thoughts can affect the confidence with which those thoughts are held (Tormala et al., 2002). Other forms of processing fluency have also been shown to affect confidence (see Koriat & Levy-Sadot, 1999, for a review). For example, Gill, Swann, and Silvera (1998) used a priming procedure to make certain kinds of impressions of a person easier to form following a paragraph about that person. Gill et al. found that when judgments were made more easily, because of priming, those judgments were also held with greater confidence. Norwick and Epley (2003) found that participants were more confident that a given statement was true when the statement was easy rather than difficult to read. Busey, Tunnilliff, Loftus, and Loftus (1995) presented participants with a series of faces to study for a later recognition test. When participants were tested under bright rather than dim viewing conditions, which made the faces easier to see, participants were more confident.

---

3 The positive effect of mere exposure is most likely when conditions favor minimal processing of the repeated stimuli (Bornstein, 1989). As stimulus processing increases, it becomes more likely that only positive items will enhance in favorability whereas negative items will decrease. When tedium or boredom sets in with excessive exposure, even positive items can be rated more negatively (Cacioppo & Petty, 1979).
in their recognition judgments. Borrowing from the mere exposure paradigm, researchers have also found various forms of repetition to increase confidence. Repeating questions, for instance, increases confidence in the answer retrieved (e.g., Hastie, Landsman, & Loftus, 1978). Similarly, repeated expression of one’s attitude increases attitude certainty, and this effect can stem from the objective ease, or accessibility, with which the attitude comes to mind (Holland, Verplanken, & van Knippenberg, 2003).

**Perceived Knowledge**

Another meta-cognition associated with cognitive fluency involves people’s perceptions of their own knowledge (see Koriat & Levy-Sadot, 1999, for a review). As but one example, Werth and Strack (2003) presented people with trivia-type questions and answers in an easy-to-read or difficult-to-read color scheme. When the colors made the question and answer easy rather than difficult to read, participants were more likely to assume they knew the answer all along. This *feeling of knowing*, it turns out, can have important implications for thought and behavior. Under some conditions, the feeling of knowing has a positive impact on information processing. The greater one’s feeling of knowing of an elusive item in memory, the more time one will spend searching for that item before giving up (e.g., Costermans et al., 1992; Koriat, 1993; Nelson & Narens, 1990; Yzerbyt, Lories, & Dardenne, 1998). Under other conditions, though, perceived knowledge has a negative impact on information processing. In particular, the more one thinks one knows about a topic, the less likely one is to seek new information on that topic (Radecki & Jaccard, 1995). These effects are particularly intriguing given that the correlation between perceived and actual knowledge tends to be quite low (e.g., Glenberg, Wilkinson, & Epstein, 1982; Radecki & Jaccard, 1995; Krosnick, Boninger, et al., 1993).

**Caveat**

As the preceding review indicates, much of the research on processing fluency suggests that the experience of ease is positive. In fact, some recent work explicitly concludes that fluency is by its very nature a pleasurable experience (for a review, see Reber, Schwarz, & Winkielman, 2004; Schwarz, 2004). Based on most of the research we have reviewed, this conclusion makes sense. Indeed, easy-to-process stimuli tend to be rated more favorably than difficult-to-process stimuli (e.g., Bornstein, 1989; Jacoby, 1983; Lee & Aaker, 2004). Easy-to-process pictures have been shown to elicit greater positive affect than do hard-to-process pictures (Winkielman & Cacioppo, 2001). Easy-to-generate thoughts or arguments, whether positive or negative, tend to be viewed as more valid (Tormala et al., 2002), frequent (Schwarz, 1998), or generally diagnostic of what one thinks. In short, processing fluency has typically been viewed as inherently positive in some way.

However, recent research has determined that the meaning of ease is malleable (e.g., Winkielman and Schwarz, 2001), and changing the meaning of ease can modify the normal ease of retrieval effect. In one study, for instance, Briñol, Petty, and Tormala (2006) reported that under some conditions people can be induced to view easy-to-generate thoughts as less rather than more diagnostic of what they think. Briñol et al. (2006) manipulated the meaning of ease by leading some participants to believe that cognitive ease was a positive sign of mental functioning, whereas difficulty was a negative sign. Other participants received the exact opposite information. When ease was described as positive, Briñol and colleagues replicated the usual ease effect—participants were more favorable toward an issue after they generated an easy rather than difficult number of arguments in favor of it. When ease was described as negative, however, this effect was reversed—participants were less reliant on their thoughts when those thoughts were easy rather than difficult to generate. In other words, although there may be a default tendency to think easy-to-generate thoughts are more plentiful, valid or desirable, this tendency can be changed (see also Freitas, Azizian, Travers, & Berry, 2005; Schwarz, 2004). The implications of this effect for other fluency effects have yet to be fully explored, but this initial finding suggests that the impact of the cognitive experience of ease on social judgment may ultimately prove more complex than is typically presumed.

**Attitudes and Persuasion**

One of the first areas in social psychology to be concerned with meta-cognition was that on attitudes and persuasion. In particular, researchers have been interested in various meta-cognitive properties of attitudes as indicators of the attitude’s strength, though the early work on this topic did not link explicitly to the “meta-cognitive” moniker. *Attitude strength* refers to the extent to which an attitude persists over time, guides behaviors and other judgments, and is resistant to influence attempts (Krosnick & Petty, 1995). Wegener and colleagues (1995) noted that for many aspects of attitude strength, there were both objective and subjective indicators. That is, for virtually every postulated objective indicator of an attitude’s strength such as the actual speed with which an attitude comes to mind (accessibility; see Fazio, 1995), or the amount of information people can generate regarding their attitudes (knowledge; see Wood et al., 1995), or the amount of thinking they have done about their attitude (Petty, Haugtvedt, & Smith, 1995), there is a parallel measure of the *perceived* ease of attitude access or amount of knowledge or thought. However, there are some subjective perceptions, such as attitude certainty and importance, for which there are no objective counterparts.

In an influential paper, Bassili (1996) referred to
the subjective perceptions regarding one’s attitudes as meta-attitudinal indicators and contrasted them with the more operative or objective indicators that tapped more directly into attitude structure or process. Notably, Bassili was rather critical of meta-attitudinal features of attitudes arguing that they are typically not represented with the attitude object nor closely related to the factors that determine attitudes. He argued that because people are often not aware of the processes leading to their judgments (Nisbett & Wilson, 1977), reports of how much knowledge was behind one’s attitude or how quickly one’s attitude would come to mind would not relate very well to operative measures of these constructs. Furthermore, he argued that “second order judgment (i.e., judgment about judgment) requires a level of intrapsychic awareness that people seldom have” (p. 640; see also Roese & Olson, 1994).

After conducting two studies comparing meta to operative indicators of strength in their abilities to predict attitude stability and resistance to change, he concluded that “operative measures have more predictive validity than meta-attitudinal measures” (p. 651). Although this conclusion is consistent with the data he reports for indices collapsed across various meta and operative measures, it is important to note that his data also show that one meta measure in particular – attitude certainty – either did as well as or outperformed the operative measures in predicting attitude consequences.

Thus, our point of view on the utility of meta-cognitive measures of attitude strength is a bit different. Consistent with Bassili (1996), we think it is unlikely that there are a large number of subjective assessments that are directly linked to attitudes. That is, it may not be the case that perceptions of accessibility, knowledge, and so forth are linked to most attitudes. In contrast to Bassili, however, we argue that many attitudes at least have a validity, certainty, or confidence tag (Petty, 2006; Petty et al., 2006). Indeed, considerable research is consistent with the idea that when people come to disbelieve or have some doubt in an idea (i.e., hold the idea with less than complete certainty), this idea can be marked with a negation or doubt tag (Clark & Chase, 1972; Gilbert, Tafarodi, & Malone, 1993; Mayo, Schul, & Burnstein, 2003).

We review research in four topic areas. First we examine research that assesses meta-cognitive strength features of attitudes and how these features affect various consequences. Second we discuss research that examines how meta-perceptions of an attitude might be affected by accepting or rejecting a persuasive message. Next, we discuss research in which meta-perceptions of one’s attitude-relevant thoughts determine persuasion. Finally, we briefly mention work regarding meta-cognition about the processes of persuasion.

Meta-cognitive Properties of Attitudes: Certainty and Others

As noted earlier, there are many potential thoughts people could have about their attitudes such as how quickly they come to mind, how many others share their view, and so forth (Wegener et al, 1995). Nevertheless, several perceptions of attitudes have received the most attention and we discuss these next.

Attitude Certainty

The most studied meta-cognitive aspect of attitudes and the one of most longstanding interest (e.g., Allport, 1924) is the certainty or confidence with which an attitude is held. Certainty generally refers to a sense of validity concerning one’s attitudes (Gross, Holtz, & Miller, 1995). Although certainty naturally covaries with extremity (i.e., people tend to feel more certain as their attitudes deviate from neutrality; e.g., Raden, 1989), certainty and extremity are conceptually distinct such that a person can have high certainty in the validity of a neutral attitude, or express an extreme attitude with low confidence. A number of determinants of attitude certainty have been examined. People tend to be more certain of their attitudes when they are based on direct experience (e.g., Fazio & Zanna, 1981), when they come to mind easily (Haddock, Rothman, Reber, & Schwarz, 1999), when others agree with the attitude (e.g., McGarty, Turner, Oakes, & Haslam, 1993; Visser & Mirabile, 2004), and when people have done much prior thinking about the attitude object (Abelson, 1988) unless that thinking arouses conflicting thoughts (Liberman & Chaiken, 1991). Some individual differences have also been related to attitude certainty (e.g., uncertainty orientation; Sorrentino, Raynor, Zubeck, & Short, 1990; dogmatism; Palmer & Kalin, 1991), though it is not clear if the certainty effects of these individual differences are independent of attitude extremity. Attitude certainty, at least with respect to political issues, tends to increase from young adulthood to middle age, but then declines sharply as one becomes elderly (Visser & Krosnick, 1988). Given the declines in memory accessibility as one ages (Varhaeghen & Salthouse, 1997), it would be interesting if declines in attitude accessibility are partially responsible for the declines in certainty in one’s later years.

Attitude certainty has been associated with a number of important attitude-relevant outcomes. In particular, attitudes held with greater certainty are more resistant to change (e.g., Kiesler & Kiesler, 1969), persistent in the absence of a persuasive attack (Bassili, 1996), and more predictive of behavior (Fazio & Zanna, 1978) than attitudes about which there is doubt. In fact, attitudes may have to reach a certain level of certainty before action is initiated (Gerard & Orive, 1987). Certain attitudes may be more resistant to change because certainty induces a confirmatory information seeking style (e.g., Swann & Ely, 1984), and certain people are more likely to assume that others agree with them (Marks & Miller, 1985). Gross et al. (1995) suggest that it is
useful to distinguish “true confidence” in one’s attitude from “compensatory confidence.” The former is based on knowledge or social support whereas the latter actually reflects an absence of confidence. We suggest that the latter might be revealed in low confidence on an implicit measure. Gross et al. argue that those with compensatory confidence may be most likely to project their attitudes onto others, and respond to counterattitudinal messages with a feeling of threat rather than challenge (see Blascovich, 1992).

**Importance, Ambivalence, and Knowledge**

Other meta-cognitive features of attitudes that have been examined with some frequency include attitude importance, ambivalence, and knowledge. Although these constructs are often determined by the same factors as attitude certainty and produce similar effects, the accumulated evidence suggests that there are sufficiently different antecedents and consequences to treat them as distinct (see Visser, Bizer, & Krosnick, 2006, for a comprehensive review). For example, in one study (Visser et al., 2003), despite an overall positive correlation between importance and certainty on the issue of global warming, it was found that as people were exposed to more media reports on global warming, the importance of global warming increased, but certainty in one’s attitudes decreased.

**Attitude importance** has been defined as the extent to which people attach significance to their attitude and care about it (Krosnick, 1988). Because importance is viewed as something attached to the attitude (Boninger et al., 1995), it is appropriately viewed as a meta-cognitive feature. However, in much research on attitude importance, what is typically measured is the perceived importance of the attitude object rather than the attitude itself (e.g., how important is the issue of global warming to you personally?); Bizer & Krosnick, 2001, italics added; see also, Krosnick, Boninger, Chuang, Berent, & Carnot, 1993; Holbrook et al., 2005; Visser, Krosnick, & Simmons, 2003). Because of this, the construct as operationalized is closely related to the construct of issue involvement which concerns the personal importance people attach to particular attitude objects and issues (e.g., Borgida et al., 1995; Petty & Cacioppo, 1990; Petty, Cacioppo, & Haugtvedt, 1992). Furthermore, when manipulations of attitude importance have been attempted, they too have been closely related (if not identical) to manipulations of issue involvement (i.e., manipulating the personal relevance of the issue; e.g., Bizer & Krosnick, 2001; Boninger, Krosnick, & Berent, 1995). To add further confusion, some researchers have measured attitude importance (e.g., “I consider my attitude toward smoking to be important,”) but have referred to it as issue importance (van Harreveld & van der Pligt, 2004).

At the conceptual level, it seems likely that attitude importance and issue importance would often go together (see Boninger et al., 1995). Nevertheless it also seems quite easy to distinguish between attitude and issue importance in some circumstances. For example, a person might think that the issue of capital punishment is very important, but care little about what particular attitudinal position is adopted with respect to it. On the other hand, another person might be committed to adopting particular opinions that agree with their ingroup, but the issues themselves are of little consequence. In an early paper, Zimbardo (1960) distinguished the involvement people can have with an issue and the involvement with one’s particular attitudinal response. The more recent conflation of attitude and issue importance is potentially misleading if the constructs are conceptually distinct and can induce different psychological consequences, at least in some situations. For example, Petty and Cacioppo (1990) argued that the increased importance of any particular attitudinal issue would lead to enhanced information seeking and processing with respect to the issue, whereas the increased importance of any particular attitudinal position (i.e., caring about whether you favored or opposed a topic), was more likely to lead to a biased seeking and processing style.

Recognizing a distinction between the importance of the attitude issue (or object) and the particular attitudinal position may help to clarify what appears to be surprising findings in the literature. For example, Bassili (1996) found that attitude importance was positively rather than negatively related to attitude change as would be expected if importance tapped into attitude strength. However, because the measure of attitude importance he used actually assessed the importance of the issue rather than the importance of the person’s attitudinal position, this positive relation makes sense if increased importance of the issue led people to pay more attention to messages about the issue and they changed in response to processing these messages (Petty & Cacioppo, 1979). Indeed, rated issue importance has been associated with enhanced information seeking activities (e.g., Visser et al., 2003).

Similar points might be made about other traditional attitude strength assessments. For instance, although researchers have referred to attitude ambivalence, the actual assessment of ambivalence, both objective and subjective, concerns whether or not people actually have (e.g., Kaplan, 1972) or perceive themselves
to have thoughts about the attitude object or issue that are mixed in valence or one sided (e.g., “I do not find myself feeling torn between the two sides of the issue of capital punishment;” Newby-Clark, McGregor, & Zanna, 2002, italics added; see also Connor, Povey, Sparks, James, & Shepherd, 2003; Priester & Petty, 1996; Tourangeau, Rasinski, Bradburn, & D’Andrade, 1989). Thus, what has been studied in the literature might more properly be called objective and subjective issue ambivalence in that the person’s thoughts are mixed with respect to the attitude object rather than the attitudinal position per se. Although it has not been studied, it is presumably possible for people to be ambivalent about their attitudinal positions per se, but this would require assessment of whether people had both positive and negative thoughts about their attitude rather than about the issue. It seems quite possible for a person to be rather ambivalent about an issue (i.e., there are some good and bad things about capital punishment), but quite unambivalent with respect to the particular attitude position held (i.e., my thoughts about being neutral are all positive; I do not feel ambivalent about having an ambivalent attitude). Similarly, people might have little ambivalence with respect to the attitude object (i.e., all thoughts about oneself are positive), but recognize that there are both good and bad things about being so favorable (e.g., others might think you are competent but conceited). This might well explain why ambivalence and certainty have been only weakly correlated. That is, ambivalence has been assessed with respect to the attitude object or issue whereas certainty has been about the attitude itself.

Finally, it is worth noting that the same distinction can be made regarding attitude knowledge. Again, what has typically been assessed has been perceived knowledge about the issue under consideration (see Davidson, 1995; Wood, Rhodes, & Biek, 1995) rather than the attitude per se. The distinction in this case has to do with whether people are being asked to make a subjective assessment of what they know about an issue (e.g., capital punishment), versus what they know about a particular attitudinal position (e.g., strongly opposing capital punishment). Although these two constructs are likely to share much variance, it is also easy to imagine divergence. For example, people could feel that they have a lot of information about the attitude issue in general, but relatively little in support of the particular position they favor. Or, they might think they have much information in support of their position, but relatively little information overall since they are completely ignorant of the other side. The more overall information they believe they have (on both sides), the more certain they might be in the validity of the attitude, whereas the more information they believe they have in support of a particular side, the more extreme their attitude might be.

The effects of perceived knowledge have paralleled the effects of amounts of actual knowledge (e.g., see Johnson, 1994; Wood, et al., 1995). First, high perceptions of knowledge lead attitudes to be more predictive of behavior (see Davidson, 1995, for a review). In addition, people who believe they are already well informed on an issue are less likely to seek additional information on that topic, especially if they are also high in the need for cognitive closure. But, if those high in perceived knowledge are provided with information, they are more likely to think about it, differentiating strong from weak arguments (see Kruglanski et al., 2005), and are less reliant on issue-irrelevant cues (Ellis, 1996). Kruglanski and colleagues (in press) note that one’s perception of one’s own knowledge is an important factor to consider in understanding source credibility effects. That is, they argue that it is the gap between one’s own perceived knowledge and that of the source that determines the source’s impact (see also, Ellis & Kruglanski, 1992).

### Attitude Certainty and Persuasion

Having reviewed some definitional issues with respect to meta-cognitive features of attitudes, we now turn to how the most studied meta-cognitive strength feature, attitude certainty, can be affected by receipt of a persuasive message. The voluminous research on persuasion has naturally focused on whether or not an attitude changes in valence or extremity as a result of message exposure. More recent research, however, has studied how the certainty with which people hold their attitudes can be affected following persuasion or resistance.

#### Confidence in Old and New Attitudes as a Result of Exposure to Persuasive Messages

First, consider a situation in which people receive a persuasive message and do not change their attitudes in response to it. Traditional analyses assume that when attitudes fail to change, the persuasive message has been ineffective. Yet, Tormala and Petty (2002) argued that when people resist a persuasive message they sometimes contemplate their own resistance and as a result, the confidence with which they hold their initial attitude can change. In particular, in several studies, Tormala and Petty showed that when people believe they have effectively resisted a strong persuasive message, certainty in their original attitudes is increased. When people believe they have resisted a weak message, certainty does not increase because people are uncertain as to whether
they would have been able to resist a stronger message.  

Subsequent research (Tormala & Petty, 2004a) showed that this meta-cognitive effect was only present among individuals who were prone to engaging in considerable amounts of thought (i.e., those high in need for cognition; Cacioppo & Petty, 1982) or in situations that fostered high amounts of thinking (e.g., a topic high in personal relevance; Petty & Cacioppo, 1979). Furthermore, just as certainty in one’s original attitude can be increased when people believe they have resisted a message with strong arguments, so too can their certainty be increased when they believe they have resisted a message from an expert (Tormala & Petty, 2004b). Resisting a message from a non-expert is not very diagnostic with respect to one’s attitude because one does not know if the attitude could have survived an attack from someone more knowledgeable. What if people are trying to counterargue a message, but are unsuccessful in doing so? In a series of studies, Rucker and Petty (2004) found that people who process a message with an intent to find fault, but are unsuccessful in counterarguing, become more convinced in the validity of their new attitude than people who processed the same message in a more objective manner. Rucker and Petty argued that when the arguments in a message are very strong, both people who are processing objectively and those trying to find fault will realize that there are many positives to the proposal. Only people who are trying to find fault and fail, however, will realize that there are few negatives to the proposal. Because these individuals will realize that they considered both sides of the issue (i.e., positives and negatives), they will have more confidence in their new attitudes. This logic implies that if the message arguments are very weak, it is people who are trying to be favorable, and fail, who will have more confidence in their old attitudes than people who are being objective. This is because everyone will realize that the proposal has many flaws (due to the very weak arguments), but only people who were trying to be favorable who will realize that there are no positives. Because these individuals will have considered both sides of the issue, confidence in their attitudes will be enhanced (Rucker & Petty, 2005).

6 In a somewhat similar vein, when people refuse to engage in an attitude consistent behavior for a high incentive, they become even more extreme in the direction of their initial opinions than when they refuse for a lower incentive. Presumably, people reason that if they didn’t go along and were given a high incentive to do so, their original attitude must really be valid. Increasing attitude extremity is one way to assert the validity of the attitude (Darley & Cooper, 1972). If attitudes did not become more extreme as a result of the manipulation, it would be reasonable to expect that confidence in the original attitude would have increased via similar attributional reasoning.

7 Of course, if people counterargue their own position rather than an external message, they can lose confidence in their attitude; Koriat et al., 1980).

The studies we just reviewed suggest quite clearly that following an attempt to resist persuasion, people sometimes reflect on their resistance and its meaning for their attitudes. In the examples just provided, certainty in either an old or a new attitude was increased as a result of reflection upon one’s attempted resistance. These outcomes are depicted in Figure 1, along with situations in which confidence in either old or new attitudes can be decreased as a result of reflection upon one’s resistance. For example, the figure suggests, and recent research confirms, that if people resist changing to a message but believe that their counterarguing attempt was flawed in some way (Tormala, Clarkson, & Petty, in press), or that they resisted by illegitimate means (e.g., rejecting a minority source; see Tormala, Petty, & DeSensi, in press), certainty in one’s attitude can be decreased (see Petty, Tormala, & Rucker, 2004, for additional discussion).

**A Meta-cognitive Approach to Attitude Change**

The notion that people can hold both old and new attitudes with varying degrees of confidence has important implications for an analysis of what happens when attitudes change. Traditional models of persuasion hold that when attitudes change, the old attitude just disappears or is incorporated into the new attitude (e.g., see Anderson, 1971). More recent constructivist perspectives on attitudes (e.g., Schwarz & Bohner, 2001) hold that attitudes are newly constructed on each occasion when they are needed, and thus there is no residue from prior attitudes. In contrast, our meta-cognitive model of attitudes, called the PAST (Past Attitudes are Still There) model when applied to attitude change situations, incorporates the notion that both old
and new attitudes can coexist with differing levels of confidence (Petty & Briñol, 2006; Petty et al., 2003, 2006).

In particular, consistent with various “negation tag” models of judgment (e.g., Gilbert et al., 1993; Mayo et al., 2003), the PAST model holds that when people shift from one attitude to another, the old attitude can be tagged as invalid or low in confidence (see Figure 2 for a person who starts out liking chocolate cake, but then comes to dislike it after reading a message about its fat content). An important implication of this model is that if people do not retrieve the associated invalidity tag that is linked to a rejected attitude, then the two opposing evaluations (old and new) could be jointly activated producing an ambivalent-like state. Notably, this ambivalence would not be explicit because when thinking carefully, people would recognize that they only truly endorse one side. However, when not thinking carefully, activation of both evaluations (without invalidity tag) could produce a state of implicit ambivalence.

In order to examine this possibility, in a series of studies, Petty and colleagues (2006) created initial attitudes in participants and then changed them for one group, and reinforced them for another group. Thus, at time 2, there were two groups of individuals who felt positively or negatively about some attitude object. The only difference was that one group had always felt positively or negatively whereas the other group used to feel the opposite way (i.e., their attitudes were changed). Petty et al reasoned that if people who used to feel differently experienced some implicit ambivalence, they should engage in greater information processing with respect to a message relevant to the attitude, just as do individuals who experience explicit ambivalence (e.g., Maio, Bell, & Esses, 1996). Consistent with this idea, when people had a new attitude that conflicted with their old one, their attitudes were more influenced by message quality than when there was no conflict with an old attitude. In research on the PAST model to date, confidence in old and new attitudes was not assessed. Yet, the results from this research are consistent with the notion that attitudes have validity tags that depending on their activation, can influence peoples’ reactions to attitude-relevant objects.

Thought Certainty and Persuasion: Self-validation Effects

In addition to considering the consequences for attitude confidence of attempts at attitude change, persuasion researchers have also begun to study how meta-cognitive processes might contribute to attitude change itself. Perhaps the best example of this comes from the self-validation hypothesis (Petty, Briñol, & Tormala, 2002). The key idea of this hypothesis is that just as attitude confidence is an important determinant of which attitudes predict behavior, thought confidence is an important determinant of which thoughts predict attitudes.

Considerable research has demonstrated that when people care about an issue or are motivated and able to think for other reasons, the number and valence of thoughts they have in response to a message determines the extent of attitude change (see Eagly & Chaiken, 1993; Petty, Ostrom, & Brock, 1981; Petty & Wegener, 1998, for reviews). The self-validation hypothesis suggests that in addition to number and valence, it is also important to consider thought confidence. Research has shown that measuring the confidence people have in their thoughts to a persuasive message enhances prediction of the attitudes that are formed over and above a consideration of the valence and number of thoughts (e.g., Petty et al., 2002). Furthermore, direct manipulations of thought confidence have a similar impact. In one study, for instance, following exposure to a message containing strong or weak arguments and a typical thought listing task (see Cacioppo, Petty, & Harkins, 1981), people were
asked to think about situations in which they had felt confident or doubtful in their thinking (see Petty et al., 2002). Those who generated instances of confidence became more certain of the validity of their thoughts than those who generated instances of doubt. Furthermore, this confidence led to greater persuasion when the message arguments were strong and to less persuasion when the arguments were weak. This is because confidence led people to rely on the favorable thoughts generated to the strong arguments and the unfavorable thoughts they generated to the weak arguments. Individuals who were induced to doubt the validity of their thoughts were less reliant on them in forming attitudes even though the number and valence of thoughts was the same as those induced to feel confidence.

The self-validation framework provides a new explanation for how variables can impact attitudes. For example, prior research on head nodding had assumed that nodding one’s head in a vertical (versus horizontal manner) produced more positive attitudes either because vertical head nodding biased thinking in a favorable direction (Wells & Petty, 1980) or because head nodding served as a relatively simple affective cue (Tom et al., 1991). The self-validation hypothesis suggested another possibility – that just as vertical head movements from others give us confidence in what we are saying, our own vertical head movements could give us confidence in what we are thinking. In a series of studies, Briñol and Petty (2003) found that head movements affected the confidence people had in their thoughts, and thereby had an impact on attitudes. Thus, when thoughts to a message were mostly favorable, vertical movements led to more confidence in the favorable thoughts generated and to more favorable attitudes than when horizontal movements were made. When thoughts were mostly unfavorable, however, vertical movements led to more confidence in the unfavorable thoughts generated and to less favorable attitudes than when horizontal movements were made.

The self-validation hypothesis proposes a new role that variables can play in persuasion situations in the context of the Elaboration Likelihood Model of persuasion. That is, in addition to serving as cues, arguments, or affecting the number and valence of thoughts that come to mind (see Petty & Cacioppo, 1986), variables can also influence the confidence people have in their thoughts and thus whether they are relied upon in forming attitudes. The self-validation hypothesis therefore provides a new mechanism by which classic persuasion variables can have an impact on attitudes.

For example, prior research has shown that the expertise of the source and a person’s mood can serve in multiple roles (e.g., argument, cue, affecting processing; see Petty & Wegener, 1998, for a review). Research on the self-validation hypothesis has shown that expertise and mood can also affect thought confidence. Source expertise presumably affects thought confidence because expert sources are more likely to present accurate information. If people can be more confident in the information presented by experts (Kaufman, Stasson, & Hart, 1999), then they can also be more confident in their thoughts to this information (see also, Kruglanski, et al., 2005). The mood prediction follows directly from the finding that people feel more confident when in some moods (e.g., happy) than in others (e.g., sad; Tiedens & Linton, 2001). In relevant research, Briñol, Petty, and Tormala (2004) showed that when the likelihood of thinking is high, and people learn of the source’s expertise after they have processed the message, source expertise influences thought confidence (see also Tormala et al., 2005). Briñol, Petty, and Barden, (2005) have shown that in similar situations, people are more confident in their thoughts when placed in a happy rather than a sad mood following a message. If people are more confident in their thoughts to a message when they learn the message is from an expert and when they are in a happy mood, this means that source expertise and a happy mood can increase persuasion relative to a non-expert and a sad mood when the arguments are strong, but decrease persuasion when the message arguments are weak.

Finally, we remind readers that the self-validation hypothesis provides an alternative explanation for some cognitive fluency effects. For example, Tormala and colleagues (2002) found that people were more reliant on their thoughts when they were asked to generate an easy rather than a difficult number of them. In addition to the classic ease of retrieval effect, perceptions of fluency and thought confidence might be involved in other persuasion phenomena. Consider the voluminous work on matching and tailoring in persuasion. Matching refers to instances in which a message is matched to some aspect of an individual. For example, an image appeal might be presented to a person high in self-monitoring (Snyder & DeBono, 1985), or an emotional message might be presented to a person whose attitudes are based primarily on affect (e.g., Fabrigar & Petty, 1999; see Petty, Wheeler, & Bizer, 2000; Briñol & Petty, 2005, for reviews of matching work). According to a fluency account, if a message matches the person in some way, it may be easier to process. In one demonstration of this logic, regulatory fit (see Higgins, 2000) was shown to enhance processing fluency. Lee and Aaker (2004) presented promotion- and prevention-focused individuals with persuasive messages framed in terms of gains or losses. Participants reported that it was easier to process the messages when they matched (gain frame-promotion focus or loss frame-prevention focus) rather than mismatched (gain frame-prevention focus or loss frame-promotion focus) participants’ regulatory focus. Furthermore, participants were more persuaded by the messages under matched rather than mismatched
conditions. Although there are several possible explanations for this effect (see Petty et al., 2000), it may be that the processing fluency stemming from regulatory fit leads message thoughts to be held with more confidence (see also, Cesario, Grant, & Higgins, 2004). To the extent that the thoughts to the message are primarily favorable, enhanced confidence would increase persuasion. This fluency explanation for various matching effects should be examined in future research.

It is noteworthy that the self-validation findings results have been most pronounced under high thinking conditions. For example, research on head nodding (Briñol & Petty, 2003), source expertise (Briñol et al., 2004), and ease of retrieval (Tormala et al., 2002) showed that these variables affected confidence in thoughts for high but not for low need for cognition individuals, and affected confidence when issue involvement was high but not when it was low. Relatively high elaboration presumably enhances self-validation effects for at least two reasons. First, if people have few thoughts, then thought confidence will have little effect. Second, the same variables that would increase elaboration (e.g., issue importance) would also likely increase thinking about one’s thoughts. If people do not care enough to generate thoughts in the first place, they are hardly likely to care enough to think about the validity of their thoughts.

Meta-Cognition about Persuasion Processes

Earlier in this section we noted that people sometimes think about their resistance to persuasion and generate inferences about their attitudes based on this resistance. More generally, research suggests that people have developed naïve theories about the persuasion process including what kinds of strategies they would use to persuade others (e.g., Bisanz & Rule, 1989), what strategies might be effective in persuading them (Friestad & Wright, 1994, 1995), and what strategies they use to resist influence (Jacks & Cameron, 2003). However, there is little (if any) empirical work on how these naïve theories influence actual persuasion. Such research could be quite informative.

Although there is little accumulated research about how people’s chronic persuasion theories affect attitudes, some work has begun to explore how people’s on-line theories of attitude change can modify the persuasion process. For example, research by Mazursky and Schul (2000) suggests that people sometimes reflect upon their information processing strategy following receipt of a persuasive message so that they can modify it for future messages if the initial strategy proves ineffective. In two studies investigating this issue, they gave people information about a consumer product along with information about the source of the information. For half of these individuals, one piece of information provided by the source was said to be invalid, whereas the other half did not learn this. Then, all participants received another message about a new consumer product either from this source or a more credible one. Some of the decision makers were highly involved with the consumer product whereas others were not. When consumers had not received any prior invalid information, judgments of the involved participants appeared to be based on their effortful consideration of the evidence presented whereas those less involved relied on simple cues, consistent with much work on dual process models of persuasion (e.g., Chaiken, Liberman, & Eagly, 1989; Petty & Cacioppo, 1986). That is, high involvement participants took longer to make their decisions than low involvement individuals (an indication of their more deliberative processing), and low involvement individuals relied more on the credibility of the information source (see Petty, Cacioppo, & Goldman, 1981). However, when recipients had received some invalid prior information, participants appeared to change their information processing strategy. That is, now it was the low involvement individuals who took more time to evaluate the messages, and it was the high involvement individuals who relied more on the source credibility information.

The Self and Individual Differences

People can have evaluations of all sorts of attitude objects. One of the most studied is the self (Baumeister, 1998). The self includes one’s cognitive representation of oneself, composed of self-schemas and self-knowledge, as well as one’s evaluation of oneself, or self-esteem. Similar to any other attitude object, the relevant beliefs (self-concept) and evaluations (self-esteem) toward the self can be accompanied by meta-cognitions about their origin, content, evaluation, amount, and so on. This section reviews the importance of considering various meta-cognitive aspects of the self.

Like research on attitudes, most of the meta-cognitive research in the self domain has focused on the confidence dimension. That is, researchers have studied peoples’ subjective certainty about the validity of their cognitive representation of themselves and their self-evaluations. Similar to the literature in other domains, confidence typically has been measured by asking people to rate the degree to which they were certain or uncertain about their self-beliefs or their self-esteem. Across different measurement and induction techniques, research reviewed in this section reveals that it is critical to consider confidence in order to understand the functioning of the self-concept and self-esteem. In the first part of this review, we cover research examining the  

---

9 Although directly asking people to rate their certainty has been the most used procedure to assess confidence, other more indirect, techniques sometimes have been used such as the intra-individual standard deviation of self-esteem scores (see, e.g., Wright, 2001, for a review). Alternatively, these indirect measures can be considered as consequences of self-certainty.
consequences of self-concept confidence for various outcomes. Next, we describe studies dealing with self-esteem confidence and its implications for a variety of relevant areas. In the final section, we move from the study of meta-cognition regarding general beliefs and evaluations about the self to the domain of meta-cognition about more specific self-dimensions. Thus, the last section explores meta-cognition and individual differences.

Consequences of Self-concept Confidence

Similar to the literature in attitude strength (e.g., Petty & Cacioppo, 1986; Petty & Krosnick, 1995), self-beliefs that are held with greater confidence are also more stable, more resistant to change, and more predictive of behavior. For example, people who report greater certainty about their self-beliefs have been found to be more stable in their self-views (Pelham, 1991). Furthermore, Pelham and Swann (1994) showed that people are more likely to actively solicit self-consistent social feedback (i.e., to actively seek out feedback that supports their existing self-views) in domains in which they are most certain. Pelham and Swann (1994) also showed that interaction partners are more likely to confirm peoples' confident rather than doubtful self-views. In short, when people report high confidence in their self-views, they are likely to behave and to be perceived in ways that are consistent with those personal views (see also Pelham, 1991).

The confidence with which self-beliefs are held can influence not only individuals’ stability and consistency but also other important dimensions, such as resistance to change. Similar to the literature in attitude strength (e.g., Bassili, 1996; Gross et al., 1995), self-beliefs that are held with greater confidence are more difficult to change. Swann and Ely (1984) found, for instance, that people who reported being relatively uncertain about their personality (e.g., extraversion) showed more change in response to a series of leading questions (i.e., a technique that leads people to provide evidence that confirms the premises in the leading questions) than those who were certain about their self-concept. In follow-up research, Swann et al (1988) replicated this finding and showed that individuals high in self-belief certainty not only resist leading questions better than those with relatively lower certainty, but boomerang effects can also occur. That is, leading questions can cause confident people to change in a direction opposite to the leading questions when encouraged to make statements that are consistent with but more extreme than their own beliefs.

Subsequent research has shown that self-belief confidence can also influence how people resist other forms of influence. For example, Sedikides (1995) found that relative to participants who reported uncertainty about their traits, those who expressed more self-belief certainty were more resistant to the biasing influences of a mood induction (happy, sad, or neutral). This research suggests that being certain about oneself can lead to more resistance to information about both the specific beliefs about which one is certain and the more general factors (e.g., mood) capable of influencing those beliefs.

In a different line of research, Baumgardner (1990) found that certainty in self-attributes not only causes people to be more resistant and to behave in a more consistent and stable way, but it can also promote a sense of control over future outcomes, thus generating positive affect. Specifically, in one study Baumgardner (1990) manipulated confidence by providing participants with bogus feedback about their personality in which they were lead to believe that based on their previous responses the experimenter was uncertain or certain about the assessment of them. Although not empirically tested, the implied assumption in this experiment was that the confidence expressed by the experimenter influenced participants’ self-certainty. As anticipated, self-certainty led to positive self-affirmation assessed by participants’ self-rating. This suggests that the confidence with which self-beliefs are held can influence not only what people think and do, but also how they feel.

Also paralleling the literature of attitude strength, the confidence with which people hold their self-related beliefs (e.g., “I’m intelligent”) has a number of implications for behavior. In general, to the extent that individuals are certain of their self-beliefs, they are more likely to act in ways that are consistent with them. For example, a person who is certain that he is humorous and lazy is likely to choose situations that allow him to be funny and avoid those that demand being productive. In research conducted to test this idea, Setterlund and Niedenthal (1993) manipulated self-certainty by asking participants to describe three times in which they acted in a way consistent (confidence) or inconsistent (doubt) with traits previously rated as highly self-descriptive. The result was that individuals who were manipulated to feel certain about their self-concept were more likely to use the self to guide decisions in a subsequent task in which they had to choose situations that allowed them to express aspects of their identity.10

In another study, Briñol and Petty (2003) found that self-belief confidence can influence self-esteem. As part of a supposed graphology study, participants were required to think about and then write down their best or worst qualities using their dominant or non-dominant hand. Writing with the dominant hand was presumed to

10 Although the manipulation in this study was intended to influence certainty, it might have just primed consistency or inconsistency thus accounting for the results. Since a manipulation check for confidence was not provided, the issue remains unclear.
induce more confidence in the self-beliefs generated compared to the non-dominant hand. Then, participants rated the confidence in their self-beliefs and reported their self-esteem. As expected, using the non-dominant hand decreased the confidence with which people held the self-beliefs they listed. This occurred despite the fact that the actual quality of the self-beliefs did not vary across the hand conditions. As a consequence of the differential self-belief confidence, the effect of the direction of self-beliefs (best vs. worst qualities) on self-esteem was significantly greater when participants wrote their beliefs with their dominant rather than their non-dominant hand. This study demonstrated that inducing doubts about positive qualities tended to undermine self-esteem whereas inducing doubt about possessing negative qualities tended to enhance self-esteem. Furthermore, this study showed that the changes in self-esteem were mediated by changes in the certainty of the self-beliefs listed.

Finally, it is worth noting that although most of the research described in this section deals with confidence as the main meta-cognitive dimension, other meta-cognitive aspects occasionally have been explored in relation to self-beliefs. For example, Pelham (1991) found that positive self-beliefs rated as important were associated with more stability than positive self-beliefs considered relatively less important (see also Sedikides, 1995). Although this result shown for self-belief importance is similar to those found for self-belief confidence, importance and confidence are two relatively independent forms of meta-cognition (see discussion in the attitude section).

Consequences of Self-Esteem Confidence

As noted earlier, people can have confidence in their self-related beliefs as well as in their overall evaluation of the self. Similar to the outcomes described in the above section, the confidence with which people hold their self-esteem has been found to have a number of important implications for different domains.

First of all, self-esteem confidence seems to moderate a variety of well-established findings in the self-esteem literature. For example, there is ample evidence that relative to individuals with high self-esteem, those with low self-esteem feel uncomfortable with success (presumably because of its inconsistency with their negative self-evaluation) and that as a consequence further success tends to be avoided. In a pioneering study, Marecek and Mettee (1972) showed that only participants with low self-esteem who also reported relatively high self-esteem certainty avoided success. Another illustration of the moderating role of self-esteem certainty can be found in the literature on self-handicapping. Self-handicapping refers to actions oriented to inhibit (or handicap) one’s own performance in order to protect oneself from others’ potential attributions that a failure, if it occurred, was due to the lack of ability (e.g., Arkin & Oleson, 1998). In their landmark study, Berglas and Jones (1978) found that individuals who were induced to doubt their abilities (by failing to solve insoluble problems) handicapped their own performance (by choosing a supposedly inhibiting drug in anticipation of a second set of similar problems) more than individuals induced to trust their abilities (by receiving success feedback on previous soluble problems). Extending this notion to the domain of self-esteem confidence, Harris and Snyder (1986) found that participants who were uncertain of their self-esteem were more likely to self-handicap (by low preparation for an upcoming test) than those who reported higher self-esteem certainty.

As another example of the moderating role of self-esteem confidence, consider work on discrepancy reduction. Social psychological literature has clearly documented that people can simultaneously hold incompatible beliefs, attitudes, feelings, and behavioral tendencies regarding oneself and others, and that these internal discrepancies are unpleasant, and often result in negative affect and psychologically undesirable outcomes (e.g., Abelson & Ronsenberg, 1958; Heider, 1958; Higgins, 1987; Kaplan, 1972; Newcomb, 1968; Norton, 1975; Osgood & Tannenbaum, 1955; Priester & Petty, 1996). A common approach to dealing with discrepancy is enhanced thinking or information processing (e.g., Abelson, et al., 1968; Aronson, 1969; Festinger, 1957). By considering additional information, individuals may hope to gain enough information for one or the other side of the discrepancy in order to resolve or minimize the inconsistency, or at least the subjective discomfort that results from the discrepancy (e.g., Briñol, Petty, & Wheeler, 2006; Katz, Wackenhut, & Hass, 1986; Maio et al., 1996). For example, Woike and Baumgardner (1993) found that participants whose global and specific self-esteem were incongruent expressed greater interest in learning more about themselves than those whose self-worth was congruent. Importantly, however, this effect was only evident for participants who reported high confidence in their global and specific self-evaluations (see also, Marsh, 1993).

Perhaps the domain of self-consistency is the area in which self-esteem certainty has been studied most extensively. Specifically, research suggests that self-esteem certainty can help to shed light on the classic debate concerning the potentially opposing human motivations for self-verification and self-enhancement. For example, people prefer feedback that is consistent with their own views when they report being certain of those views, but they prefer positive feedback when those views are less confidently held (Pelham, 1991). In line with this finding, as noted earlier, Swann and Ely (1984) found that participants who reported being certain of their traits tended to convince others to see them as they saw themselves. When participants reported being relatively
uncertain about their traits, however, others saw them according to their own expectations. Extending this work from perceptions by strangers to perceptions by known others, Pelham and Swann (1994) found that self-views (both self-beliefs and self-esteem) matched other’s views more strongly when those self-views were reported to be relatively certain. Taken together, these findings suggest that the need to be consistent operates mostly for aspects of the self-concept that are held with more confidence (see, Wright, 2001 for a review). 11

In closing this section, it is important to consider not only the consequences of self-concept confidence, but also some of its potential antecedents (for an extensive review of antecedents and consequences of self-certainly, see, DeMarree, Petty, & Briñol, in press). Similar to the literature on attitude certainty, theory and common sense suggest that the certainty of people’s self-views is likely to be grounded in the amount of information they have about themselves as well as the consistency of this information. However, as described earlier in this section, self-concept confidence can depend on and be affected by other more transitory variables in the situation, such the hand with which self views are written (Briñol & Petty, 2003) and the behavior of the experimenter (Setterlund & Niedenthal, 1993). In addition to these factors, and given that certainty in self-views is considered to be socially desirable, it seems possible that self-concept confidence might result from other operating motives related to impression management. 12

11 Since most of the studies conducted in this domain have been correlational, the presumed directionality of some of the effects remains unclear. For example, consider the studies in which self certainty was described as leading to different forms of self-consistency, such as asking for confirmatory feedback (e.g., Pelham, 1991; Pelham & Ely, 1984; Pelham & Swann, 1994). It would be plausible to argue the opposite directionality. That is, self-belief confidence might be the product, rather than the precursor, of interpersonal congruence. This argument can also be applied to the relationship between self-esteem certainty and other concepts, such as self-esteem stability and self-esteem clarity. Future research would benefit from designs in which confidence is manipulated rather than measured (e.g., Briñol & Petty, 2003).

12 To the extent that certainty in self-views is considered to be socially desirable, those with higher impression management concerns should be more likely to report greater certainty in their self-conceptions. Also, since individuals with higher self-esteem are more likely to present themselves in a positive way than are those with lower self-esteem (e.g., Baumeister, Tice, & Hutton, 1989), it may be the case that esteem differences in confidence reflect differing impression management concerns (e.g., Wright, 2001, see also Baumgardner, 1990; Story, 2004). Indeed, some other consequences of self-concept certainty described in this chapter (e.g., stability, resistance, and prediction of behavior) might also be partially due to impression management. Furthermore, people may report feeling confident in a given self-view in order to compensate for other unrelated internal doubts (e.g., McGregor, et al., 2003; Rhodewalt & Morf, 1995).

Individual Differences

As just reviewed, people can differ in their self-concepts and self-esteem as well as the confidence with which these constructs are held. There are other differences among individuals in which certainty plays a part. The term individual differences refers to how people vary with respect to factors such as personality, motives, and abilities. Importantly, some conceptualizations of individual differences can be understood as meta-cognitions to the extent that they refer to thoughts people have about their thoughts or thought processes. Consider, for example, the need for cognition (NC; Cacioppo & Petty, 1982), which refers to stable individual differences in the tendency to engage in and enjoy effortful thought. Consistent with the schema outlined in this chapter and an analysis of the NC scale items, people’s judgments about their own NC represent an evaluation of thinking (e.g., “I find satisfaction in deliberating hard for long hours”). Although the NC scale has focused mostly on enjoyment of thinking (i.e., the evaluative component), NC also is related to the assessment of the amount of thinking (e.g., “I usually end up deliberating about issues even when they do not affect me personally”). That is, judging our own NC might imply thinking about how much we tend to think and how much we enjoy thinking. Importantly, in the NC scale, these assessments are not about any particular thoughts (as in most meta-cognition research) but are about thought processes in general.

People appear to have good insight into their own enjoyment of thinking and tendency to engage in it as the need for cognition scale has proven to be a robust predictor of a wide variety of cognitive activities (see Cacioppo, Petty, Feinstein, & Jarvis, 1996, for a review. Individuals high in NC not only tend to think more about any given attitude object (e.g., Cacioppo, Petty, & Morris, 1983), but they also devote more attention to their own thinking. As a result, high NC has been related to meta-cognitive processes described in this chapter such that individuals high in NC are more likely to evaluate their own thoughts for validity (Briñol & Petty, 2003; Briñol et al., 2004; Petty et al., 2002; Tormala et al., 2002), to engage in controlled (Martin, Seta, & Crella, 1990; see Wegener & Petty, 1997) and automatic (Petty, DeMarree, Briñol, & Horcago, 2005) bias correction processes, and to draw different meta-cognitive inferences based on their responses to persuasive messages (e.g., Tormala & Petty, 2004a, 2004b).

Need for closure (NFC; Webster & Kruglanski, 1994) is another individual difference variable that could
be understood as meta-cognitive in nature. NFC refers to people’s desire for a definitive answer on some topic as opposed to confusion and ambiguity. Thus, NFC taps into a person’s thoughts about their own thinking with some items explicitly linked to confidence in thinking (“I usually make important decisions quickly and confidently”). If a person high in NFC generates a thought that is assessed as a definitive answer to address the situation, that person should have confidence in that thought and should consider it an appropriate thought. Importantly, NFC represents a stable individual difference as well as a situationally evocable state. As a chronic dimension, the desire for definitive knowledge has been measured with the need for closure scale (for properties of the scale, see Webster & Kruglanski, 1994; see also Neuberg, West, Judice, & Thompson, 1997). In general, being high in NFC has been shown to reduce the extent of information processing, to magnify primacy effects, to increase reliance on theory-driven versus data-driven processing, and also to enhance reliance on initial anchors and primes (see Kruglanski and Webster, 1996, for a review).

There are a variety of other individual differences variables that may be relevant for meta-cognition since they refer to different aspects of mental activity, such as causal uncertainty (Weary & Edwards, 1994), field dependence (Witkin, et al., 1954), self-monitoring (Snyder, 1974), uncertainty orientation (Sorrentino & Short, 1986), need to evaluate (Jarvis & Petty, 1996), self-awareness (Carver, & Scheier, 1981), preference for consistency (Cialdini, Trost, & Newsom, 1995), resistance to persuasion (Briñol et al., 2004), and defensive confidence (Albarracin & Mitchell, 2002). Although these variables deal in one way or another with how people think about their thinking, they do so in a relatively direct way. That is, most of the items contained in the scales consist of direct statements about thinking, rather than second order thoughts such as the ones described for need for cognition. For example, the self-doubt scale (Olson, Pohlmann, Yost, Lynch, & Arkin, 2000) measures individual differences in uncertain feelings about one’s competence and ability. This scale contains items, such as “More often than not I feel unsure of my abilities.” These self-doubt judgments take place at the direct, first level of cognition and are different from, for example, the second order judgments described in the section on self-concept certainty.

A clear illustration of this distinction between people’s self-perceptions and meta-cognitions can be found in the literature of individual differences relevant to minority groups. For example, social psychologists have developed numerous measures to assess individual differences in attitudes toward many groups considered to be stigmatized in some way (see Briñol & Petty, 2005, for a review). However, there are not only individual differences in evaluations of minority groups (which constitutes a direct judgment), but also individual differences in chronic motivations to control for prejudice toward these groups (which refers to the evaluation dimension or appropriateness of the former judgment). Among this second group of measures are the Motivation to Control Prejudiced Reactions scale (Dunton & Fazio, 1997), the Internal and External Motivation to Respond without Prejudice scale (Plant & Devine, 1998), and the Humanitarianism-Egalitarianism and Protestant Ethic Scales (Katz & Hass, 1988). As described in more detail in the next section, these instruments are effective in predicting differences in public and private endorsement of stereotypes as well as motivation to correct one’s social judgments for inappropriate content.

Bias Correction

Our last domain of meta-cognition relates a number of types of thoughts about thoughts, including assessments of thought content (valence and source of the thought), evaluation, and validity. Although these various meta-cognitive assessments come into play in bias correction, the essence of correction rests in evaluation of the thought or judgment as relatively good or bad, wanted or unwanted, appropriate or inappropriate. As noted earlier, this evaluation of the thought can be distinguished from identification of the thought as being relatively positive versus negative toward the target. Thus, one can perceive the thought as good or bad, appropriate or not, regardless of whether the thought itself is a positive or a negative one (e.g., “it is inappropriate to be favorable toward criminals”). Individuals’ evaluations of their thoughts and perceptions can have sweeping effects on judgment and behavior. When a thought or perception is viewed as bad, unwanted, or inappropriate, people might try to avoid or to modify the thought or perception. People might also try to limit the effects of that thought or perception on subsequent judgments and behavior.

Thoughts or perceptions can be viewed as unwanted or inappropriate for a variety of reasons. In general, these reasons can be summarized by saying that perceptions are viewed as inappropriate or unwanted when they do not serve the perceiver’s current judgment goals (Wegener & Petty, 1997). As noted earlier, in many cases, the judgment goal is likely to be arriving at a “correct” or “accurate” view of the target (Chaiken et al., 1989; Festinger, 1954; Petty & Cacioppo, 1986). Therefore, if a thought or perception is viewed as inaccurate, people are likely to be motivated to improve those thoughts or perceptions by using one of a number of possible strategies for bias correction. Of course, many other goals and motivations are possible. People might want to view themselves as possessing generally positive qualities (see Kunda, 1990) or a specific positive quality such as good health (e.g., Ditto & Lopez, 1991).
In contrast, as reviewed earlier, people with negative self-views might be motivated to continue to view themselves negatively (e.g., Swann & Ely, 1984). In some cases, goals are more socially oriented, as when people are motivated to uphold procedural justice for courtroom defendants, even if accuracy is better served by judging the defendant as guilty (e.g., Fleming, Wegener, & Petty, 1999). In research on need for closure, people have been shown to view certain thoughts and ideas as unwanted precisely because they oppose the goal of making a quick decision (e.g., Kruglanski & Webster, 1991).

Corrections can be distinguished from the previously discussed meta-cognitions by considering the extent to which the issue of bias is salient to individuals (see Wegener & Petty, 2001; Wegener, Petty, Smoak, & Fabrigar, 2004). When bias is not salient, people may seek goal-appropriate perceptions of targets (often valid or correct perceptions). Seeking correctness would bear similarities to the promotion orientation described by Higgins (1998b). When seeking correctness, many of the previous types of meta-cognitions are likely (e.g., how confident am I in my judgment). However, to the extent that potential for bias is salient, people become more oriented toward taking steps to identify and avoid any biases at work (Wegener & Petty, 1997; similar to Higgins’ prevention orientation).

We discuss three different types of corrections guided by a sense that one’s cognition are somehow unwanted or inappropriate. These corrections are: attempts at subtracting reactions to biasing factors when forming perceptions of targets, attempts to use theories of bias in seeking to formulate appropriate (goal-consistent) perceptions of targets, and attempts at suppressing or inhibiting a particular thought from coming to mind in the first place.

**Subtraction of Reactions to Biasing Factors**

Much of the research on bias correction from the 1980s and early 1990s focused on partialling or subtraction of reactions that are viewed as inappropriate because they are inferred to be responses to some irrelevant contextual variable rather than to the target. For example, research on Martin’s (1986) set-reset model generally begins with blatant priming of responses consistent with one interpretation of an ambiguous target. Overlap in reactions to the prime and to the target result in setting, which is a default misattribution to the target of reactions that were really reactions to the prime (i.e., the context in which the target is judged). That is, the default perception is that reactions are "about the target" (Higgins, 1998a) rather than about the context. Setting is said to produce assimilation to the context, which is consistent with traditional priming effects (e.g., Higgins, Rholes, & Jones, 1977; Srull & Wyer, 1979). For example, Martin et al. (1990, Study 3) blatantly primed participants with the concept persistent or stubborn and then gave participants target materials that were ambiguous with regard to whether the target was persistent or stubborn. Consistent with the idea that “setting” to blatant primes requires relatively little thought, perceivers who were low in need for cognition (Cacioppo & Petty, 1982) assimilated judgments of the target to the primed concept.

However, when motivation and ability to think are sufficiently high, the set-reset approach suggests that perceivers reset (i.e., partial out reactions to the context). Overlap in valence between reactions to the target and reactions to the context can lead to confusion over which reactions belong to the context and to the target. Thus, attempts to partial out reactions to the context could result in some reactions to the target being misattributed to the context. When this occurs, resetting not only reduces assimilation to the context, but can result in contrast (i.e., judgments of the target that are even less like the context than if the context were not present). Consistent with this idea, perceivers high in need for cognition in the Martin et al. (1990, Study 3) research contrasted their judgments of the target away from the blatantly primed concepts.

Similarly, the inclusion-exclusion model (Schwarz & Bless, 1992a) treats inclusion of information in one's representation of the target as the default mental operation and treats exclusion of the information as requiring greater cognitive effort. Inclusion-exclusion studies typically begin with a target that can either be a superordinate category (within which a specific member of the category can be included or excluded) or a subordinate member (or subset) of a category (to which characteristics of the category as a whole can be ascribed or excluded). When the context information is subordinate to the target category, including that information in the representation of the category leads to assimilation of the category to the exemplar. For example, thinking of an extreme exemplar, such as a popular athlete, could increase perceivers’ positivity toward the athlete's gender or ethnic group (e.g., Bodenhausen, Schwarz, Bless, & Wänke, 1995; Coats & Smith, 1999; Schwarz & Bless, 1992b). Exclusion of the exemplar can reduce the assimilation (if the exemplar is simply subtracted) or can lead to contrast (if oversubtraction occurs, similar to resetting, or if the excluded exemplar serves as a standard of comparison for the category; Schwarz & Bless, 1992a). The standard of comparison could directly influence perceptions of the target (as in judgment theories like adaptation level theory, Nelson, 1964, or social judgment theory, Sherif & Hovland, 1961) or could redefine the meaning of the

---

13 Attributing reactions to the target or the context is quite similar to source monitoring in research on memory, wherein perceivers attempt to determine the cause of recollective experiences such as familiarity of a target (see Johnson & Raye, 1981).
response scale anchors (as in variable perspective theory, Ostrom & Upshaw, 1968).

If information superordinate to the target comes to mind, that superordinate information can lead to assimilation if the target is included in the superordinate category. This inclusion also sets the stage for generation of additional features of the target using the category. Similar to encountering of subordinate information, exclusion of superordinate information also leads to a decrease in assimilation or to contrast. This could be because of subtraction of features associated with the category or because of establishment of an extreme standard of comparison.

Like the set-reset approach, inclusion versus exclusion is determined, in part, by whether a particular type of information is regarded as appropriate for inclusion in the representation of the target. Most studied determinants of inclusion versus exclusion directly reflect attention to the role of categorization in context effects. For example, information that is representative of the target category should generally be included, whereas information not representative of the target category should be excluded. Lack of representativeness could be caused by factors such as temporal distance between the context and target (Strack, Schwarz, & Gschneidinger, 1985), lack of feature overlap between context and target (Herr, Sherman, & Fazio, 1983), or receipt of information from separate, potentially inconsistent sources rather than from a single source (Hilton & von Hippel, 1990). Similarly, information is more likely to be included when category width is high (e.g., asking about politicians in general after asking people about three political scandals) but excluded when category width is low (e.g., asking about a particular politician not involved in the previously rated scandals; Schwarz & Bless, 1992b). Finally, when stimuli are presented together, they are more likely to be perceived as a unit, resulting in assimilation, but when stimuli are presented sequentially, they are perceived separately, resulting in contrast (e.g., Seta, Martin, & Capehart, 1979; Wedell, Parducci, & Gieselmann, 1987).

Similar to the set-reset approach, Schwarz and Bless (1992a) also noted that exclusion of reactions can occur when participants realize that previously encountered stimuli (e.g., primes) other than the target may have created the reactions (e.g., Lombardi, Higgins, & Bargh, 1987; Strack, Schwarz, Bless, Kübler, & Wänke, 1993). Schwarz and colleagues have also shown that conversational norms can motivate people to deliberately exclude information from target categories. For example, Schwarz, Strack, and Mai (1991) asked participants about their marital satisfaction and about their general life satisfaction. When people reported their life satisfaction after their marital satisfaction, the correlation between these measures was high, presumably because people included the recently-activated information about marital satisfaction when reporting life satisfaction. However, when the questions were described as assessing two areas of life that may be important for overall well-being, the correlation between the items decreased substantially because marital satisfaction information was presumably excluded from ratings of life satisfaction (see also Strack, Martin, & Schwarz, 1988).

Attempts to partial or subtract perceptions include a variety of features that are highly meta-cognitive. People have to identify the likely sources of thoughts and reactions and have to determine whether those reactions are informative about the target as currently construed. If the thought or reaction is to be subtracted, there likely has to be some cognitive mechanism for setting the thought aside and focusing attention on the included thoughts instead. In the subtraction research conducted thus far, few measurements of such meta-cognitive mechanisms have been included. For example, one might imagine that, in order to set aside a particular reaction, some type of monitoring of that reaction might have to take place. If so, then as in the thought suppression work described shortly (e.g., Wegner, 1994), the thought or reaction that is set aside might actually become rather accessible in memory as the monitoring occurs. Similarly, few studies have measured perceptions of appropriateness or of attributions of reactions to targets versus contexts. Therefore, although many of the manipulations used in this research would seem to suggest that these types of meta-cognitions are involved, future research could benefit from more direct assessment of them.

Theory-Based Correction

An alternative view on the meta-cognitive activity of bias correction relies on peoples’ perceptions of the bias(es) at work in a given setting.14 For some time, researchers have noted that people might realize that a bias is at work and might make efforts to overcome that bias (e.g., Higgins et al., 1977; Strack, 1992; Thompson, Fong, & Rosenhan, 1981; Wyer & Budesheim, 1987). Early attention to lay beliefs or theories about biases focused on the fallibility of such perceptions. For example, Wilson and his colleagues have shown that people often believe that factors affect their perceptions even when the factors do not, and people often believe that they can resist influences that they cannot (e.g., Nisbett & Wilson, 1977; Wilson & Brekke, 1994; Wilson, Houston, & Meyers, 1998).

Despite the potential inaccuracy of beliefs about bias, social perceivers may use such perceptions in attempts to avoid bias. The Flexible Correction Model is

---

a theory of bias correction based on social perceivers’ use of naive theories of bias (Petty & Wegener, 1993; Wegener & Petty, 1995, 1997). Research guided by this approach has shown that people correct their judgments in different directions when they hold theories of opposite biases (e.g., Wegener & Petty, 1995; Wegener, Petty, & Dunn, 1998), even when those opposite theories of bias are for different people perceiving the same context and target (Wegener & Petty, 1995). People correct for biases they believe exist, even if there is no real bias (e.g., Wegener & Petty, 1995). This can then create the opposite bias, as when corrections for perceived negativity toward the dislikable source of a persuasive message leads that source to be more persuasive than a likable source (Petty, Wegener, & White, 1998; see also Schul & Goren, 1997).

Corrections for perceived bias also mean that people sometimes correct primarily for one bias (the one most salient or for which clear beliefs exist), even if other biases truly influence perceptions of that same target. For example, Sczesny and Kühnen (2004) showed that people believe that gender can bias judgments of leadership qualifications but do not hold similar beliefs about physical features associated with masculine versus feminine appearance. When encountering mock application materials (including photos), research participants showed different corrections for gender versus physical features. When cognitive load was high (and meta-cognitive activity was likely curtailed), participants were more likely to employ male than female applicants and people who possessed masculine rather than feminine features. When cognitive load was low, however, research participants overcorrected effects of gender (the bias for which relevant naive theories existed), such that female applicants were viewed more favorably than male applicants. In contrast, the biasing impact of masculine versus feminine appearance (for which no naive theories existed) remained unchanged.

Research on theory-based correction also illustrates the breadth of domains in which corrections for bias occur. Studies of theory-based correction have been conducted in domains as diverse as affect and judgment (e.g., Berkowitz, Jaffee, Jo, & Troccoli, 2000; DeSteno, Petty, Wegener, & Rucker, 2000), impression formation (Iseb & Wyer, 1999; Stapel, Martin, & Schwarz, 1998), stereotyping (Lepore & Brown, 2002; Strack & Mussweiler, 2001), persuasion (Petty et al., 1998; Schul & Goren, 1997), recognition memory (Förster & Strack, 1998), and courtroom judgment (Thompson et al., 1981; Wegener, Kerr, Petty, & Fleming, 2000).

Consistent with the idea that on-line metacognition is most likely when motivation and ability to think are high, most studies of theory-based correction that have manipulated or measured motivation or ability to think have shown greater theory-based correction with high levels of thinking (e.g., DeSteno et al., 2000; Sczesny and Kühnen, 2004). There may be important exceptions to this general pattern, however. For example, if certain corrections are performed repeatedly, they may become less effortful (see Wegener & Petty, 1997; cf., Glaser & Banaji, 1999), even automatic. In one study, for instance, Maddux, Barden, Brewer, and Petty (2005) assessed the automatic evaluations that Whites had of Black versus White targets in particular contexts. The White participants varied in their motivation to control racial prejudice as assessed with a scale developed by Dunton and Fazio (1997). People who are high in motivation to control prejudice are presumably highly practiced in controlling any prejudiced reactions they might be feeling. When people high and low in motivation to control prejudice evaluated Blacks versus Whites in innocuous settings (e.g., church, garden), neither group showed much prejudice replicating prior research (see Barden, Maddux, Petty, & Brewer, 2004; Wittenbrink, Park, & Judd, 2001). However, when the same Black and White targets were evaluated in settings that might suggest anti-Black bias (e.g., jail, dingy factory), those low in motivation to control prejudice showed an anti-Black bias, but those high in motivation to control prejudice showed a significant pro-Black bias, consistent with the idea that they were (over) correcting their judgments. Because the attitude measure tapped automatic evaluative responding (see Fazio et al., 1986), this research is consistent with the idea that highly practiced corrections can be executed automatically in certain contexts.

Some biases might be so salient or obvious that people would adjust their ratings without much additional consideration of the target. This might occur in studies of the sleeper effect in persuasion when a discounting cue says that the previous message was false and, in fact, the opposite position has stronger support (e.g., Gruder, Cook, Hennigan, Flay, Alessis, & Halamaj, 1978). Such strong discounting cues are necessary for production of the sleeper effect (see Kumkale & Albarracin, 2004). From a theory-based correction point of view, strong discounting cues might produce a sleeper effect because they alert message recipients to an obvious bias that brings about an initial adjustment of ratings without theory-guided “reprocessing” of the initial information (after all, why go back to pay attention to the false information). If initial processing of the information was high and the message is encountered prior to the discounting cue (two other requirements for the sleeper effect, see Kumkale & Albarracin, 2004; Petty & Cacioppo, 1986), the relatively nonthoughtful correction only temporarily shifts ratings, whereas attitudes based on initial thinking about the information are more likely to persist over time. In other circumstances, however, more thoughtful corrections could lead to persisting “corrected” views (Wegener & Petty, 1997), which would undermine the sleeper effect (see Priester,
Wegener, Petty, & Fabrigar, 1999, for additional discussion). Whether corrections are accompanied by considerable or little thought is a fertile area for future research.

There might also be circumstances in which high levels of initial thinking make biases harder to identify and to correct (Petty & Wegener, 1993). For example, high levels of elaboration should lead to a great deal of integration of perceptions with existing knowledge structures (Petty & Cacioppo, 1986). Depending on the type of knowledge available about (or related to) the target, the highly integrated view of the target may seem justified by the existing information. In other words, if the reactions to the target seem to be an accurate reflection of available information, it may be that the people would view their opinions of the target as relatively appropriate and unbiased. Of course, this would undermine any need to correct the perception of the target (see Wegener, Clark, & Petty, in press; see also Schul & Burnstein, 1985). High levels of integration could also spread the bias across many disparate perceptions. This might also make identification of bias and correction more difficult.

Although significant research supports the possibility of corrections based on peoples’ perceptions of bias, much work remains to be done. Future research will likely address more directly the meta-cognitive mechanisms at work when people spontaneously identify potential biases. It could be that people use accessible or salient theories to guide searches for potential bias, but there could be a variety of additional cues to bias such as a mismatch between past and present perceptions of the target, matches between the valence of a salient situational factor and the current perceptions, or others (see Wegener et al., 2001, for additional discussion).

**Thought Suppression**

On some level, if it were possible, keeping biased thoughts from coming to mind at all would be the ideal way to avoid bias. Yet, research on thought suppression attempts have suggested limited utility in this strategy. For example, Wegener and his colleagues have found that many types of thoughts become hyperaccessible after initial attempts to suppress them (e.g., Wegner, 1994; Wegner & Erber, 1992). In one study, Macrae, Bodenhausen, Milne, and Jetten (1994) asked participants to write a paragraph about a day in the life of a skinhead either with or without instructions to avoid stereotypic thoughts about the target. Later, a lexical decision task showed that the stereotype of skinheads was more accessible (i.e., people were faster to recognize stereotype consistent words as words) when the initial essay had been written while attempting to suppress stereotypic thoughts. According to Wegener and colleagues, this ironic hyperaccessibility of the suppressed content comes from a monitoring process that scans conscious thought for the presence of the unwanted material. As this process checks for the to-be-suppressed material, it repeatedly activates the concept in memory.

Consistent with on-line meta-cognitive activity being more likely when motivation and ability to think are high, attempting to suppress thoughts is an attention-consuming activity (Macrae, Bodenhausen, Milne, & Wheeler, 1996), and ability to suppress is undermined with time pressure (e.g., Wegner, Erber, & Bowman, 1993; described in Wegner, 1994). Thought suppression may become easier and more effective in avoiding rebound with practice (Kelly & Kahn, 1994; Monteith, Sherman, & Devine, 1998; cf., Smith, 1994; Wegner, 1994). Some reasons for this may include that people who practice suppressing particular thoughts might also have more replacement thoughts (i.e., thoughts that distract one from the to-be-suppressed concept) at the ready (e.g., Monteith et al., 1998; Wegner, Schneider, Carter, & White, 1987) and more practiced individuals may become less likely to have the unwanted thoughts come to mind in the first place (e.g., Fazio, Jackson, Dunton, & Williams, 1995; Lepore & Brown, 1997).

Although there is sufficient evidence that attempts to suppress thoughts can make the thoughts more accessible at a later point in time, it is important to note that many studies of thought suppression allow for other types of correction in addition to or instead of suppression per se. For example, Wyer, Sherman, and Stroessner (2000) showed that suppression of stereotypes toward African Americans resulted in application of primed stereotype-consistent concepts to a race-unspecified target, but not to an African American target. Moreover, this restriction in use of primed stereotype-consistent material for a race-identified target only occurred when people had sufficient cognitive capacity to engage in the meta-cognitive work of correction. Although it is possible that people re-suppressed the primed concepts when encountering the African American target, it seems more likely that research participants were using one of the other corrective mechanisms to limit the impact of the primed concepts on judgments of the target. For example, Dunton and Fazio (1997) noted that motivation to control prejudice might create overcorrections for automatically activated negative reactions through the types of theory-based correction proposed in the FCM ( Petty & Wegener, 1993; Wegener & Petty, 1997; see also Maddux et al., 2005). When the race of the target was unspecified in the Wyer et al. (2000) research, perceivers might not have realized that their perceptions of the target could be biased by the activated reactions, thereby opening the door to the bias (see Monteith et al., 1998; Szesny & Kühnen, 2004; Strack & Hannover, 1996; Wegener & Petty, 1997).

In addition to suppression of thoughts per se, one could also talk about suppression of judgments or behaviors. That is, rather than focusing on keeping a
thought from coming to mind, people might focus on making sure a thought does not become an action (cf., Monteith et al., 1998). In the social judgeability approach, Yzerbyt and his colleagues have pointed out that people might withhold judgments when they do not feel justified in making a judgment. This might often occur when people perceive the amount of information about the target to be insufficient for forming an accurate view of the target (e.g., Darley & Gross, 1983; Yzerbyt, Leyens, & Corneille, 1998). Interestingly, this caution in judgment can be overcome by people believing they have received individuating information about the target even if they really did not (Yzerbyt, Schadron, Leyens, & Rocher, 1994). Therefore, the perception of having received sufficient information can lead to greater willingness to use stereotypes in formulating judgments. The willingness to make stereotype-consistent judgments is likely due, in part, to perceiving the judgment as more appropriate or less biased when based on an information base that seems sufficient. The social judgeability research also provides a potential bridge between thought evaluation and perceptions of confidence/validity. As noted earlier, perceived lack of accuracy can lead to perceptions of thoughts or reactions as unwanted or inappropriate. In the Yzerbyt et al. (1994) research, people express greater confidence in their judgments when they believe they have received individuating information. That is, when people lack confidence in the validity of their reactions, they are unwilling to use those reactions for judgments, but when people possess greater confidence, they are willing to use those reactions.

Even when people prefer not to be prejudiced toward a target, inability to inhibit the use of stereotype-related information can influence the likelihood of making judgments biased by the target's group membership. For example, elderly perceivers, who have trouble inhibiting a variety of types of thoughts about available information (Hasher & Zacks, 1988) also are more likely to provide stereotypic judgments, even if they are instructed to avoid race-related information and even when older perceivers profess greater motivation to avoid prejudiced responding (von Hippel, Silver, & Lynch, 2000).

Comparison of Strategies

To date, there is precious little research directly comparing subtraction, suppression, and theory-based corrections. In an initial attempt, Strack and Mussweiler (2001) compared recomputation of judgments (i.e., setting aside of biased information and formulation of perceptions based on remaining information, Strack & Hannover, 1996; Strack & Mussweiler, 2001; cf., Schwarz & Clore, 1983) with adjustment of responses (generally consistent with theory-based corrections or subtraction). Strack and Mussweiler (2001) designed studies so that recomputation would lead to more stereotypic judgments (because available information was highly stereotype consistent), but adjustment would lead judgments to be less stereotypic. When participants received a brief resume about a hypothetical job applicant, an instruction asking people not to be influenced by gender led to less stereotypic ratings of the applicant. This could have been due to theory-based correction or to subtraction effects. However, when a large number of stereotype-consistent behavioral episodes were presented about the target, the same correction instruction led to an increase in stereotypic ratings (consistent with recomputation). In a follow-up study, Strack and Mussweiler (2001) provided participants with the large number of behavioral episodes, but varied whether perceivers were given a goal of forming an accurate judgment or of being fair and not using gender to make judgments. Similar to the earlier study, the accuracy goal led to judgments that were consistent with the presented information. However, the fairness goal led to judgments that were less consistent with gender stereotypes.

In two studies, Dove, Wegener, and Petty (2001, 2003) used a similar design, but made the same amount of information either available or unavailable just prior to judgment. When the information was available, judgments were consistent with recomputation. However, when information was unavailable, judgments were consistent with theory-based corrections. It is possible that recomputation is more likely when perceivers form their opinions in an on-line rather than memory-based fashion (Hastie & Park, 1986). Yet, there might also be a variety of settings in which memory-based recomputation is feasible. When participants received a conditional correction instruction (i.e., they were asked to correct if they perceived anything as biasing them; see Stapel et al., 1998), both the recomputation and theory-based correction patterns were more pronounced for people high in need for cognition (Cacioppo & Petty, 1982).

How to interpret such effects remains a matter for discussion. Strack and Mussweiler (2001) interpreted their results as signifying a preference for recomputation when both options were open to social perceivers. However, some type of adjustment might happen more often, with recomputation limited to situations in which social perceivers see recomputation as feasible and reliable (Dove et al., 2001, 2003). At any rate, comparisons among the correction processes are just beginning. There are likely to be a host of moderators of such outcomes. For example, it seems unlikely that accuracy-based instructions would always result in recomputation when information is available about the target. This should depend on factors such as the accessibility and salience of the theory of bias as well as the salience of the biasing agent itself (Wegener et al., 2001). Undoubtedly, future work will include additional
comparisons of the correction processes, including studies of moderators of when each process dominates.

Basic Principles and New Directions

In the present chapter, social psychology’s major research findings on meta-cognition were described. Consistent with most prior literature (e.g., Jost et al., 1998), we defined meta-cognition as second order thoughts, or thoughts about our primary thoughts or thought processes. That indeed constitutes a basic principle in human cognition:

**Principle 1:** There is primary and secondary cognition. Primary thoughts are those that occur at the direct level of cognition, involving initial associations. Following a primary thought, people can also generate other thoughts that occur at a second level which involve reflections on the first level thoughts or the process that generated these thoughts.

Perhaps the most important conclusion of the work we reviewed is that in general, second order cognition has an impact on first order cognition. This is summarized in principle 2.

**Principle 2:** Second order cognition can magnify, attenuate, or even reverse first order cognition.

When increasing confidence in an attitude causes it to exert a larger impact on behavior than it did previously, meta-cognitive factors are exerting a magnifying effect on one’s attitudes (see Fazio & Zanna, 1978). When people shaking their heads while listening to a message rely on their thoughts less than they would have if they had not been shaking, meta-cognitive factors are exerting an attenuating factor on one’s thoughts (see Briñol & Petty, 2003). When people want to control for their perceived biases and then show an opposite bias, metacognitive processes are reversing the effects of first order cognition (see Petty, Wegener, & White, 1998; Wegner, 1994).

In reviewing the relevant literature, we argued that the same categories that have proven effective for classifying primary thoughts could also be used to categorize meta-cognitive thoughts. This is summarized in principle 3.

**Principle 3:** Second order thoughts can be coded into the same categories that have already proven effective for classifying primary thoughts, such as target, evaluation, number, and confidence.

By grouping the many specific meta-cognitions into meaningful categories, we aimed to provide a useful guide to organize and facilitate access to key findings across diverse literatures in social psychology.

The main areas of social psychological research in which meta-cognition has been examined most extensively are: (a) memory and cognitive fluency, (b) attitudes and persuasion, (c) the self and individual differences, and (d) bias and correction processes. Although different categories of meta-cognition were examined in these different areas, it is clear that across areas, the most studied dimension has been the confidence one has in one’s thoughts and judgments. As reviewed, confidence has proven to be relatively independent of accuracy, and has been shown to be affected by a wide variety of situational and individual variables.

One area that has not been explored much in prior research concerns the various bases of confidence. For example, in one study Haugtvedt and Petty (1992) found that although both high and low need for cognition individuals changed their attitudes in response to a message from a high credible source, and developed equivalent levels of confidence in these attitudes, the confidence stemmed from different sources. In particular, for those high in need for cognition, confidence increased as the number of message arguments they could recall increased. For those low in need for cognition, however, confidence was tied to the perceived credibility of the source. Though speculative at the moment, it seems quite plausible to argue that different bases of confidence might lead to different outcomes (see also, Gross et al., 1995). For example, it could be that some bases of confidence are more likely to be consequential than others.

In addition to examining the content underlying confidence, it would also be important to examine the processes leading to confidence judgments. For example, just as primary judgments can be based on much or little thought (Kruglanski & Thompson, 1999; Petty & Cacioppo, 1986), so too can second order cognition. That is, people’s judgments of confidence can be based on a careful consideration of the reasons why they should be confident, or they can result from a simple inference based on shaking one’s head. To the extent that second order cognition follows the same principles as primary cognition, one would expect thoughtful confidence to be more long lasting and consequential than confidence that stems from simple cues and inferences.

Finally, it would be worth studying whether meta-cognitive judgments are formed on-line or are retrieved from memory (Hastie & Park, 1986). Just as primary judgments can operate in either way, so too do we propose that this holds for meta-cognitive judgments. To date, there is no research conclusively demonstrating that meta-cognitive tags (e.g., certainty, importance) have any structural basis in memory, though some research is certainly consistent with this possibility. These considerations lead us to our first speculative postulate that is worthy of research attention.
Principle 4: The content and process bases of meta-cognitive judgments are likely to be as consequential as are the bases of primary cognition.

In addition to examining confidence (and other meta-cognitive constructs) based on different content and process considerations, a potential avenue for future research would be to assess various meta-cognitive properties with implicit measures. This raises the possibility that just as some researchers have argued that primary cognition can be held at explicit and implicit levels, so too might this be the case with secondary cognition. That is, a person might have high confidence in some judgment at the explicit level, but have low confidence at the implicit level. Furthermore, this implicit uncertainty could potentially guide thinking and behavior (e.g., Briñol, Petty, & Wheeler, 2005). As described earlier in this chapter, for example, when attitudes are changed, people may sometimes have high confidence in their new attitudes at the explicit level, but have low confidence at the implicit level due to the conflict with the old attitude (Petty, et al., 2006). This leads to our second speculative principle.

Principle 5: Although explicit meta-cognitive activity is generally more likely to take place when people have the motivation and ability to attend to and interpret their own cognitive experiences, meta-cognition might also operate outside awareness with important consequences for social judgment and behavior.

Indeed, the possibility of assessing mental constructs at the implicit level might open the door for a new generation of research relevant to meta-cognition. For instance, a common characteristic of most of the research covered in our review is that meta-cognitive activity is more likely to take place when people have the motivation and ability to attend to and interpret their own cognitive experience. Research on bias correction has shown, for example, that in order to correct for mental biases, people need to identify the likely source of their thoughts, determine whether those reactions are informative, and to spend time and attention trying to modify, suppress, substitute and/or correct those thoughts (Petty & Wegener, 1993; Wilson & Brekke, 1994). In contrast to all these metal activities that require extensive cognitive effort, recent research has shown that correction processes can also occur automatically if the correction is a highly practiced one (Maddux, et al., 2005). Future research should explore what variables moderate the relationship between explicit and implicit meta-cognition. In sum, although most of the research covered in this chapter has focused on the power of explicit meta-cognition to modify the impact of explicit primary cognition, both primary and secondary thoughts may also operate outside awareness with important consequences for social judgment and behavior.

References


Krosnick, J. A. (1988). The role of attitude importance in social


and attitude change. In D. Granberg & G. Sarup (Eds.), Social judgment and intergroup relations: Essays in honor of Muzafer Sherif (pp. 147-174). New York: Springer-Verlag.


Strack, F. (1992). The different routes to social judgments: Experiential versus informational based strategies. In L.L. Martin and A. Tesser (Eds.), *The Construction of*


